AD-A065 771 STANFORD UNIV CALIF SYSTEMS OPTIMIZATION LAB F/G 12/1

FUNDAMENTALS OF A CONTINUOUS TIME SIMPLEX METHOD.(U)

DEC 78 A F PEROLD N00014-75-C-0267

UNCLASSIFIED SOL-78-26 NL

1 OF 2

AD
A065771

Systems
Optimization
Laboratory

0

LEVEL

AD A0 65771

DDC FILE COPY

Department of Operations Research
Stanford University
Stanford, CA 94305

SYSTEMS OPTIMIZATION LABORATORY
DEPARTMENT OF OPERATIONS RESEARCH
Stanford University
Stanford, California
94305

9 Technical rept.

LEVEL II

408 765

6

FUNDAMENTALS OF A CONTINUOUS TIME SIMPLEX METHOD

by

10 André F. Perold

12 153 p.

TECHNICAL REPORT 14 SOL-78-26

11 ~~December~~ Dec 1978

15 N00014-75-C-0267, EY-76-S-03-0326-PA-18

Research and reproduction of this report were partially supported by the Department of Energy Contract EY-76-S-03-0326 PA #18; the Office of Naval Research Contract N00014-75-C-0267; and the National Science Foundation Grants MCS76-81259 A01; MCS76-20019 A01 and ENG77-06761 A01.

408 765

xlt

TABLE OF CONTENTS

TABLE OF CONTENTS Cont.

CHAPTER 1

INTRODUCTION

The general continuous linear program is formulated as follows:

$$\text{minimize} \qquad \int_0^T c(t)\, x(t)dt$$

subject to

$$B(t)\, x(t) + \int_0^t K(t,s)\, x(s)ds = b(t)$$

$$x(t) \geq 0, \qquad t \in [0,T]$$

where $c(t)$, $b(t)$ and $B(t)$, $K(t,s)$ are given and are real vectors and matrices respectively.

A special case of this is a linear optimal control problem with constraints on both the state and control variables:

$$\text{minimize} \qquad \int_0^T \{c(t)\, x(t) + d(t)\, u(t)\}dt$$

$$\begin{aligned} \text{subject to} \qquad \frac{d}{dt}\, x(t) &= A(t)\, x(t) + B(t)\, u(t) + a(t) \\ 0 &= C(t)\, x(t) + D(t)\, u(t) + b(t) \\ x(t) &\geq 0, \quad u(t) \geq 0, \quad t \in [0,T] \\ x(0) &\ \text{given.} \end{aligned}$$

The motivation for studying such problems is the following:

(i) They are widely applicable to many real world situations as intertemporal economic models of, say, investment and planning (e.g. [3] and [7]), and occur frequently in engineering applications (e.g. [24]).

(ii) They are closely related to their discrete time counterparts which, formulated as block triangular linear programs, are costly to solve. Computational experience with such linear programming models shows that they often require unusually many simplex iterations [1], [7],[1] and also become very large in size. On the other hand, many small continuous time problems have been observed to have nice mathematical properties, and exact solutions have been easily obtainable by hand [12], [18]. This suggests that a thorough understanding of continuous linear programs may not only result in methods that can solve them directly and efficiently on a computer, but also result in new improved methods for solving the discrete time formulations.

(iii) Many continuous time problems are inherently numerically unstable when solved in discrete time as linear programs. Indeed, one can construct examples which are easily solved in continuous time, but which, when discretized with time step ϵ, have

[1] This seems to be attributable to the observed persistence property of these models [8], [21], i.e. similar type activities persist in the basis for several consecutive time periods. In the continuous case this can be interpreted as activities remaining positive over intervals of time.

corresponding linear programming bases with condition numbers proportional to ϵ^{1-m}, where m is the number of rows in the continuous time problem. In such cases, some other solution technique becomes a necessity for any appreciably small ϵ.

Continuous linear programs have been largely studied as linear programs in a function space with the emphasis on generalizing the simple but powerful results of linear programming. They were first considered in 1953 by Bellman [2],[3] who established a weak duality result which, as a sufficient condition for optimality, could be used to obtain optimal solutions by "good guesswork." Since then work has been mainly in two areas: strong duality theorems and computational methods.

The first strong duality theorem was established by Tyndall [26] in 1965, and subsequently strengthened by a number of authors, e.g. Grinold [14] and Levine and Pomerol [19]. These authors imposed algebraic conditions on the coefficients defining the problem so as to directly generalize the strong duality theorem of linear programming [6] to the case of continuous linear programs in the space of bounded measurable functions. Another strong duality theorem requiring a Slater condition to be satisfied was obtained by Hager and Mitter [15] for a variant of the above optimal control formulation.

The dominant theme on the computational side has been the attempt to generalize the simplex method to a function space. In 1964 Dantzig [9] showed that the control theory formulation with no state variable constraints could be solved using the Dantzig-Wolfe decomposition principle. Work on the general problem was done first by Lehman [18]

in 1954 and then pursued by Drews, Hartberger and Segers [12] in 1970. Their central idea was to mirror the revised simplex method step-by-step in continuous time. Thus they would begin with a "basic feasible solution" which, roughly speaking is a solution whose positive values are uniquely determined once the remaining ones are held fixed at zero. The first step would be to compute relative "prices" backwards in time by requiring complementary slackness conditions to hold. These prices would then be used to identify some variable that is currently at zero on some subinterval of $[0,T]$, and which when increased over an interval would yield an improved solution. This increase would be made in such a way that the new solution was again a "basic feasible solution."

Teren [24] in 1977 developed a similar algorithm for the optimal control formulation allowing, in addition, for end point constraints. However, his regularity assumptions are very restrictive.

The role of the above works, [18], [24] and especially [12], has been to demonstrate that the concepts and steps of the simplex method can be directly generalized to a function space setting. However, many questions are left unanswered, and some of their suggested constructions need to be made more precise if they are to be developed with mathematical rigor.

In this dissertation, we consider some of the foundational aspects of a theory of a continuous time simplex method in the same spirit as [12] and [18].

The first task we address is that of obtaining a useful characterization of extreme point solutions. This is achieved, in Chapter 2, in the case of constant coefficients and in the class of so-called

right analytic solutions. The main result there is that there is a one-to-one correspondence between extreme points and solutions satisfying certain full rank conditions, in a manner reminiscent of the analogous characterization in linear programming, and also the bang-bang principle of optimal control theory. This characterization is used to define a "basic feasible solution" and is the foundation of the proposed continuous time simplex method in Chapter 4. Examples are also presented to show that this kind of characterization cannot hold for the general (time varying coefficients) case. This gives some indication that it might be meaningless to look for a general continuous time simplex method.

In Chapter 3 we motivate the approach to be taken in Chapter 4. This is done by first reviewing the important steps in the simplex method, and then analyzing an example in both continuous and discrete time.

Chapter 4 is chiefly concerned with the problem of how to move from one basic feasible solution to an improved one. The first aspect considered is the statement of the optimality conditions. It is well known, e.g. [12], [18], that the dual problem need not have an optimal solution in any space of real valued functions even though the primal may be bounded and possess a unique analytic optimal solution. While the existence of simple examples to this effect motivated the search for conditions on the coefficients where this would not occur, e.g. [14], [26], it also simultaneously yielded the realization that any continuous time simplex method would have to be able to accommodate these cases, [12], [18]. Indeed, working in the larger function space of distributions[1] which contains the Dirac

[1] See Appendix C.

δ functional and its higher order derivatives allowed these dual problems to attain their optima. However, interpreting a dual infeasibility (related to a statement of non-optimality) in the space of distributions requires some care, and this is an aspect ignored in [12] and [18].

We here take the view that it is reasonable to assume that the primal problem is bounded. This means that the need to work with distributions arises only in the dual. We then formulate a new dual problem with all quantities being <u>real valued</u>. This dual is shown to be equivalent to a formulation in the space of distributions that allows only finitely many occurrences of the δ and its derivatives, but which nevertheless can be studied in an entirely distribution free context. The usual weak duality theorem is established and sufficient optimality conditions deduced.

Next is presented a new construction which, under certain non-degeneracy and regularity assumptions, will begin with a given extreme point solution and from there continuously trace out a path of extreme points along which descent is obtained for at least a short while. It differs fundamentally from the usual basis change in linear programming in that we are here simultaneously increasing nonbasic variables having both negative and positive reduced costs, but so that the overall result still gives descent. The approach taken here was inspired by Dantzig [10] who suggested that dual shadow prices could be used to find which nonbasic variables to introduce into the basic set because their values would dominate any changes in the objective due to corrections in the location of the breakpoints in order to maintain feasibility. This thesis gives conditions (and proofs) for which his conjecture holds.

Chapter 5 contains the conclusions and mention of some of the many remaining unanswered questions. The appendices contain examples, background material on distributions and Laplace transforms, and also technical lemmas, some of which are of independent interest. Appendix D in particular contains an interesting existence and uniqueness result for a set of simultaneous nonlinear equations, the solutions to which are the breakpoints in one of the basis change constructions.

We conclude this introduction with a remark on the optimal control problem formulated earlier. Because of the presence of mixed control and state variable constraints, the problem is very different in character from and substantially more difficult to solve than the usual control problem having the state variables unrestricted. See e.g. [13]. Indeed, the distinction between state and control variables becomes artificial, in general yielding no additional insight. When it is reformulated as a continuous linear program,[1] the state and control variables are lumped together in a single vector. Thus, rather than thinking of choosing a control to steer the system along a certain trajectory, we instead take the combinatorial approach of choosing the active constraint sets from amongst both the state and control variables.

1.1. Notation

Let $\underline{n}$ denote $\{1,2,\ldots,n\}$.

For $\beta \subseteq \underline{n}$, let $|\beta|$ be the cardinality of β.

For an $m \times n$ matrix, A, and $\alpha \subseteq \underline{m}$, $\beta \subseteq \underline{n}$, let $A_{\alpha\cdot}$ denote the submatrix of A whose rows are indexed by α; let $A_{\cdot\beta}$ denote

[1]By integrating both sides of the system equation.

the submatrix of A whose columns are indexed by β; let a_i denote the ith column of A, and a_{ij} its (i,j)th element. All matrices will be denoted by upper case letters.

Vectors will be denoted by lower case letters, with the distinction between row and column vectors being omitted when the meaning is clear from the context. When necessary the transpose symbol 'T' will be used to indicate a row vector, b^T, for some column vector b. The components of a vector b will be denoted by b_i. The distinction between this and the ith column of the matrix B will be clear from the context.

e_p will denote the pth column of the identity matrix, and e will be used to denote a vector of ones.

$g^{(i)}(t)$ will denote the ith derivative of g evaluated at t. $g^{(i)}(t+)$ will denote the ith right derivative of g, and similarly for $g^{(i)}(t-)$.

$\|\cdot\|$ will denote the Euclidean norm.

If $I = (t',t'')$ is an open interval then $\tilde{I}$ will denote the interval $[t',t'')$.

R and $\mathbb{C}$ will denote the reals and complexes respectively.

$L_\infty[0,T]$ will denote the space of all real valued, Lebesgue measurable, essentially bounded functions on $[0,T]$. $L_1[0,T]$ will denote the space of real valued Lebesgue integrable functions on $[0,T]$. When the time interval is clear from the context, these will be written as L_∞ and L_1 respectively.

C^∞ will denote all infinitely differentiable functions from $R \to R$.

$1_+(t) = 0$ for $t < 0$, $= 1$ for $t \geq 0$.

Chapters are numbered from 1 through 5, with sections and subsections denoted by 4.1, 4.2.1, etc. Appendices are numbered A through E with subsections likewise denoted by C.1, C.2, etc. Equations, lemmas, propositions and theorems are all part of the same numbering system within a chapter. 4(25), 4.2(25), 4.2.1(25) all refer to the same equation 25 in Chapter 4, and will be used when referenced from another chapter.

[6] will denote reference number 6 in the bibliography.

CHAPTER 2

ON EXTREME POINTS AND THEIR CHARACTERIZATION AS BASIC FEASIBLE SOLUTIONS

The simplex method of linear programming is based partly on the following fundamental results [6].

(i) If $f:R^n \to R$ is linear, and f is bounded below on

$$P = \{x \in R^n : Ax = b,\ x \geq 0\}$$

where $A \in R^{m \times n}$ and $b \in R^m$, then f attains its infimum on P at an extreme point of P.

(ii) $x \in P$ is an extreme point of P if and only if $A_{\cdot\beta}$ has full column rank, where $\beta = \{i : x_i > 0\}$.

Because of the characterization in (ii), extreme points of polyhedral sets are also called 'basic feasible solutions.' Note that there can be at most m positive components in any basic feasible solution.

In this chapter we shall investigate to what extent similar statements may be made in the context of continuous linear programs. We shall begin with the general case, and then specialize various aspects of the problem.

The general problem under consideration is

Minimize

$$f(x) = \int_0^T c(t)\ x(t)dt$$

(1)

Subject to

$$B(t)x(t) + \int_0^t K(t,s)\ x(s)ds = b(t)$$

$$x(t) \geq 0$$

where $K(t,s)$, $B(t) \in R^{m \times n}$, $x(t)$, $c(t) \in R^n$, and $b(t) \in R^m$ for all s, $t \in [0,T]$, $s \leq t$.

It will be useful to write the constraints in operator shorthand:

$$Ax = b, \qquad x \geq 0$$

where A maps the function x to another function denoted by Ax and defined by

$$(Ax)(t) = B(t)x(t) + \int_0^t K(t,s)\ x(s)\ ds.$$

In order to complete the problem definition we need to specify the spaces in which the variables and coefficients lie, e.g. L_∞. This will be done when required.

For a given space X and constraint coefficients $B(\cdot)$, $K(\cdot,\cdot)$ and $b(\cdot)$ define

$$P(X) = \{x \in X^n : Ax = b,\ x \geq 0\}$$

where X^n denotes all n-vectors of elements in X. $P(X)$ is easily seen to be convex, and will be assumed nonempty.

We shall identify any two functions that are equal a.e., and likewise require the constraints to be satisfied a.e. In particular, if x is an extreme point of $P(X)$ and $y = x$, a.e., then y is also an extreme point of $P(X)$.

2.1. Extreme Points as Optimal Solutions

The result that would be most desirable is the following: if X is a given space of 'nice' functions, and if a continuous linear functional f is bounded below on $P(X)$ then f attains its infimum on $P(X)$ and moreover does so at an extreme point of $P(X)$. However, without severe pre-conditions on the coefficients defining the problem, the only case when this appears possible is when $X = L_\infty$ and $P(L_\infty)$ is bounded. One complicating factor is that while the notion of extreme points is purely algebraic, one seems to require heavy topological machinery to establish merely their existence.

(2) Theorem. If the components of $c(\cdot)$, $b(\cdot)$, $B(\cdot)$ and $K(\cdot,\cdot)$ are all in L_∞, and there is an $M > 0$ such that $x \in P(L_\infty) \Rightarrow \|x(t)\| \leq M$ a.e. then $f(\cdot)$ attains its infimum at an extreme point of $P(L_\infty)$.

The proof of this result will follow immediately from the following well-known lemma once the compactness of $P(L_\infty)$ is established in a suitable topology.

(3) Lemma. Let Y be a locally convex Hausdorff space and let $f: Y \to R$ be concave. If $Q \subseteq Y$ is compact and f is lower semi-continuous on Q then f attains its infimum on Q at an extreme point of Q.

Proof. See [16, p. 74]. □

(4) Lemma: Compactness of $P(L_\infty)$

Under the conditions of Theorem (2)

(i) $P(L_\infty)$ is weakly compact as a subset of L_1^n.

(ii) $P(L_\infty)$ is weak * compact.[1]

Proof.

(i) See [14, p. 40].

(ii) The proof of this is in the same spirit as that of (i) and will be omitted. □

Proof of Theorem (2).

Let Y be the space L_1^n with the weak topology. By Lemma (4) $P(L_\infty)$ is compact in Y. Since $c \in L_\infty^n$, f is continuous on Y. Noting that Y is a locally convex Hausdorff space, we can apply Lemma (3), and the proof is complete. □

---

[1]For interest sake we state both compactness results even though only one is required.

Remark. The same proof is valid if we choose Y to be L_∞^n with the weak * topology.

In practice we would like to have optimal solutions that are more manageable than general measurable solutions, for example piecewise analytic solutions having only finitely many breakpoints. Theorem (2) unfortunately is the best statement available and it is still an open question whether or not it can be improved upon even in very special cases. Even if we know that the optimum has, say, a piecewise analytic solution, there is no guarantee that there is a piecewise analytic extreme point solution. However, there is a motivation for continuing the study of extreme points in more useful spaces, given by the following simple result.

(5) **Proposition.** If $f:X^n \to R$ is concave and x is the unique minimizer of f over $P(X)$, then x is an extreme point of $P(X)$.

Proof. This follows immediately from the concavity of f and the definition of an extreme point. □

2.2. Extreme Points as Basic Solutions

Let us begin by reviewing the proof of the characterization of extreme points in R^n given at the beginning of this chapter.

We have a given $x \in R^n$ satisfying

$$Ax = b, \qquad x \geq 0.$$

The necessary and sufficient condition for x to be an extreme point is that $A_{\cdot\beta}$ has full column rank, where $\beta = \{i : x_i > 0\}$.

To show the sufficiency, we write $x = \lambda y + (1-\lambda)z$ for $\lambda \in (0,1)$ and some y and z satisfying the constraints, and then show that $x = y = x$. Set $\alpha = \{i : x_i = 0\}$. $x_\alpha = 0$, $y \geq 0$ and $z \geq 0$ implies $y_\alpha = z_\alpha = 0$. Therefore $A_{\cdot\beta}y_\beta = A_{\cdot\beta}z_\beta = A_{\cdot\beta}x_\beta = b$. Since $A_{\cdot\beta}$ has full column rank these equations have a unique solution, and we are done.

To show the necessity, assume that $A_{\cdot\beta}$ does not have full column rank. Then there exists a $y_\beta \neq 0$ such that $A_{\cdot\beta}y_\beta = 0$. Setting $y_\alpha = 0$ and noting that $x_\beta > 0$, there is a $\theta > 0$ such that $x + \theta y \geq 0$, $x - \theta y \geq 0$. Hence writing $x = \frac{1}{2}(x + \theta y) + \frac{1}{2}(x - \theta y)$ shows that x is not an extreme point.

It is precisely these two steps that we shall mirror in the continuous case, namely

(i) being able to solve uniquely for the positive components when the remaining ones are held at zero, and

(ii) being able to perturb the positive components to either side when they are not uniquely determined.

Before continuing, we remark that in order to obtain a similar characterization in a function space we shall have to severely restrict the class of admissible solutions as well as the constraint coefficients. The reason for this is that the constraints in (1) can have extreme points for which no characterization of the above form is possible.

Example B(1) is such a case. Also, for computational purposes it would be preferable if we could solve locally for $x(t)$ at time t, without any considerations of the future. In Example B(1) the values of x on $[0,1)$ are determined by constraints that hold over the interval $[1,4)$. In Example B(2) the slope of x at 0, and hence the whole solution, is determined only at time T by the restriction $x(T) = 0$.

As in the case of linear programming we are here aiming at the following restatement of the continuous linear program (1): Choose a partition of $[0,T]$ into time intervals $\{I_j\}$ and an associated partition of the variables $\{(\alpha_j,\beta_j)\}$ such that with $x_{\alpha_j} = 0$ on I_j, the remaining variables x_{β_j} are uniquely determined, and yield the optimal solution. By our comments in the preceding paragraph, we moreover want the value of $x(t)$ to be determined only by our choice of which variables are to be held at zero on $[0,t]$, and not by what happens to the right of t.

In the remainder of this chapter we shall work with the space of right analytic functions, defined below.

(6) Definition. A function $g:[0,T] \to R$ will be called right analytic if for each $t \in [0,T)$, there is an $\epsilon > 0$ and an analytic function $h:(t-\epsilon,t+\epsilon) \to R$ such that $g(s) = h(s)\ \forall\, s \in [t,t+\epsilon)$.

We shall let $\mathcal{A}_r$ denote the space of bounded right analytic functions on $[0,T]$. The required properties of these functions are established in Appendix A. Our motivation for choosing the class $\mathcal{A}_r$ is that it seems to be the largest class for which the local uniqueness result, Proposition (9) below, can be established.

The following is a key result in the subsequent analysis.

(7) Lemma. Let $g:[0,T] \to R^n$ have right analytic components $g_i(\cdot)$, $i = 1,\ldots,n$. Then there exists a (possibly infinite) disjoint family of open intervals, $\{I_j\}$, such that $\cup \tilde{I}_j = [0,T)$,[1] and such that for each interval I_j, each g_i satisfies

(i) g_i is analytic on I_j

(ii) either $|g_i| > 0$ on I_j or $g_i = 0$ on I_j.

Proof. See Appendix A. $\square$

For a given $x \in P(\mathcal{C}_r)$ and its associated partition $\{I_j\}$, let

$$\alpha_j = \{i : x_i = 0 \text{ on } I_j\} ,$$

and

$$\beta_j = \{i : x_i > 0 \text{ on } I_j\} .$$

Let t_j' and t_j'' denote the endpoints of I_j. Then using the constraints (1), the equation for x_{β_j} on I_j becomes

$$(8) \qquad B_{\cdot\beta_j}(t)\, x_{\beta_j}(t) + \int_{t_j'}^{t} K_{\cdot\beta_j}(t,s) x_{\cdot\beta_j}(s) ds = d^j(t)$$

for all $t \in [t_j', t_j'')$, where $d^j(t) = b(t) - \int_0^{t_j'} K(t,s)x(s)ds$.

[1]If $I = (t',t'')$ is an open interval then $\tilde{I}$ denotes $[t',t'')$.

(9) Proposition. Let $x \in P(\mathcal{C}_r)$ and let $\{I_j\}$ be the associated partition of $[0,T]$ given by Lemma (7). If equation (8) has a unique right analytic solution on each interval $\tilde{I}_j$, then x is an extreme point of $P(\mathcal{C}_r)$.

Proof. Suppose that x is not an extreme point. Then there exist $y, z \in P(\mathcal{C}_r)$ and $\lambda \in (0,1)$ such that $x = \lambda y + (1-\lambda)z$ and for some $t \in [0,T)$, $x(t) \neq y(t)$. By Lemma A.5 there exist $0 \leq r < s \leq T$ such that $x = y$ on $[0,r)$ and $x \neq y$ on (r,s). By Lemma (7) there is an interval $I_j = (t_j', t_j'')$ of the partition such that $t_j' \leq r < t_j''$. By analogy with the sufficiency argument presented in R^n, it is clear that on I_j, y_{β_j} satisfies

$$B._{\beta_j}(t)\, y_{\beta_j}(t) + \int_{t_j'}^{t} K._{\beta_j}(t,s)\, y_{\beta_j}(s)ds = e^j(t)$$

for all $t \in [t_j', t_j'']$ where

$$e^j(t) = b(t) - \int_0^{t_j'} K(t,s)\, y(s)ds \,.$$

However, since $y = x$ on $[0,r)$, it follows that $e^j(t) = d^j(t)$ on I_j. Thus y_{β_j} satisfies (8) on I_j. Since $y_{\alpha_j} = x_{\alpha_j} = 0$ on I_j, $y_{\beta_j} \neq x_{\beta_j}$ on $I_j \cap (r,s) \neq \emptyset$, contradicting the uniqueness hypothesis. $\square$

Remarks.

(i) In sum we have shown that being able to solve uniquely for the positive components locally is indeed a sufficient condition for a right analytic solution to be an extreme point. The important part in the proof played by the right analyticity was that when given $x, y \in P(\mathfrak{a}_r)$ with $x \neq y$, we could find an earliest interval, I_j, in the partition, on which $x \neq y$. An example where we cannot find a first interval is the following: Let $x(t) = [t \sin(1/t)]^+$ (i.e., positive part), and $y(t) = t$ on $[0,1]$. Clearly, in the partition of $[0,1]$ induced by x, there is no first interval on which x and y differ. Note further that the theorem is false if the solution is right analytic but we choose a partition $\{I_j\}$ that does not satisfy $[0,T) = \cup_{j=1}^{\infty} \tilde{I}_j$. This is illustrated in Example B(2).

(ii) Example B(1) shows that local uniqueness is not, in general, a necessary condition for a solution to be an extreme point.

We can now proceed to find algebraic conditions on the coefficients that ensure unique solutions to equations of the type

$$(10) \qquad D(t)\, x(t) + \int_{t'}^{t} L(t,s)\, x(s) ds = g(t), \qquad t \in [t', t'')$$

Since, by Example B(1), such conditions cannot in general also be necessary conditions for uniqueness, we shall confine ourselves to the case where the necessity has been established. This is the time invariant case, i.e. when D and L are constants.

Before doing so, we remark briefly that in the event $D(t) = I$, equation (10) is a Volterra equation of the second kind, and it is well-known that such equations always have unique solutions, provided that the coefficients $g_i(\cdot)$ and $L_{ij}(\cdot,\cdot)$ are in L_2. See for example [25, p. 10]. Thus if $D^{-1}(t)$ exists a.e. and $(D^{-1}g)_i$, $(D^{-1}L)_{ij} \in L_2$, we also obtain uniqueness. In general, however, $D(t)$ may be singular. This case has been studied by Doležal [11], and in differential equation form by Silverman [23]. However their work provides only a partial answer, and a general succinct uniqueness condition has, to our knowledge, yet to be discovered.

In the time invariant case, equation (10) reads

$$Dx(t) + L \int_{t'}^{t} x(s)ds = g(t), \qquad t \in [t',t'') . \tag{11}$$

This equation has been thoroughly studied in its more conventional differential equation form by a number of authors. See for example [5], [11], [23], and also Appendix C. The uniqueness condition we require is the following:

(12) Lemma. If the components of $g(\cdot)$ are analytic and x satisfies (11), then a necessary and sufficient condition for x to be the unique analytic solution is that there exist a scalar μ such that $\mu D + L$ has full column rank.

Proof. See [5]. $\square$

Remark. This full rank condition has several interesting interpretations. If D and L are square, then $\mu D + L$ has full rank iff $\det(\mu D + L)$ is not identically zero as a function of μ.[1] On dividing by μ and setting $\epsilon = 1/\mu$ the condition reads: $D + \epsilon L$ has full column rank for all ϵ sufficiently small. If we replace (11) by its discrete time analog using time intervals of stepsize ϵ, we obtain a block lower triangular coefficient matrix with each diagonal block being $D + \epsilon L$. Clearly this block triangular matrix has full column rank iff $D + \epsilon L$ has full column rank. Another interpretation can be made by taking Laplace transforms on both sides of (11) with dummy variable μ. The coefficient matrix of the Laplace transform of x so obtained is precisely $D + \frac{1}{\mu} L$.[2]

In order to show that this full rank condition is also a necessary condition for a solution to be an extreme point, we require the following lemma.

(13) Lemma. If there is no scalar μ such that $\mu D + L$ has full column rank, then given any $\tau > 0$, there exists a nontrivial analytic solution to the homogeneous equation

(14) $$Dx(t) + L \int_0^t x(s)ds = 0, \qquad t \geq 0$$

---

[1]Note that $\det(\mu D + L)$ is a polynomial in μ. Hence it is zero either for all μ or for at most finitely many μ.

[2]See also Appendix C.

satisfying

$$\int_0^\tau x(s)ds = 0 .$$

<u>Proof</u>. We use an argument similar to that given in [5 , p. 418]. By assumption, for each μ there exists a nonzero (constant) vector φ_μ such that $(\mu D + L)\varphi_\mu = 0$.

If D has k columns let $M > 2k$ be any integer and let $\mu_1,\ldots,\mu_M$ be any distinct scalars. Let G be the following $2k \times M$ matrix:

$$G = \begin{bmatrix} \varphi_{\mu_1} & \varphi_{\mu_2} & \cdots & \varphi_{\mu_M} \\ e^{\mu_1\tau}\varphi_{\mu_1} & e^{\mu_2\tau}\varphi_{\mu_2} & \cdots & e^{\mu_M\tau}\varphi_{\mu_M} \end{bmatrix}$$

Since $M > 2k$ the columns of G are linearly dependent. Hence there exists a nonzero vector $\eta = (\eta_1,\ldots,\eta_M)$ such that $G\eta = 0$.

Set

$$x(t) = \sum_{i=1}^{M} \eta_i\mu_i e^{\mu_i t}\varphi_{\mu_i} .$$

Then it is easily verified that x satisfies (14) and that $\int_0^\tau x(s)ds = 0$. Moreover, x is not identically zero. □

We can now obtain the main result of this chapter.

(15) Theorem. Characterization of right analytic extreme points. Let x be right analytic and satisfy

$$(16) \qquad Bx(t) + K\int_0^t x(s)ds = b(t), \qquad x(t) \geq 0,\ t \in [0,T],$$

where B and K are constant. Let $\{I_j\}$ be the associated partition of $[0,T]$ given by Lemma (7), and define

$$\alpha_j = \{i : x_i = 0 \text{ on } I_j\}$$

$$\beta_j = \{i : x_i > 0 \text{ on } I_j\}$$

Then a necessary and sufficient condition for x to be an extreme point is that for each j, there exists a scalar μ_j such that $\mu_j B_{\cdot\beta_j} + K_{\cdot\beta_j}$ has full column rank.

Proof. Sufficiency. By Lemma (12), x_{β_j} is uniquely determined on $\tilde{I}_j$. By Proposition (9) x is an extreme point.

Necessity. Suppose there is a j such that for all μ, $\mu B_{\cdot\beta_j} + K_{\cdot\beta_j}$ does not have full column rank. Since $x_{\beta_j}(t) > 0$ on I_j, there is a closed interval $[u,v] \subset I_j$ and an $\epsilon > 0$ such that $x_i(t) \geq \epsilon$ for $t \in [u,v]$ and $i \in \beta_j$. By Lemma (13) there exists a nonzero analytic $y_{\beta_j}(\cdot)$ satisfying

$$B_{\cdot\beta_j} y_{\beta_j}(t) + K_{\cdot\beta_j} \int_u^t y_{\beta_j}(s)ds = 0, \qquad t \in [u,v)$$

and

$$\int_u^v y_{\beta_j}(s)ds = 0 .$$

Rescale so that $|y_i(t)| \leq \epsilon$ for all $t \in [u,v)$ and $i \in \beta_j$. By construction of y_{β_j} in Lemma (13) this can always be done. Set $y_{\alpha_j} = 0$ on $[u,v)$ and $y = 0$ on $[0,T] \setminus [u,v)$. Then by construction y satisfies

$$By(t) + K \int_0^t y(s)ds = 0$$

and

$$x(t) + y(t) \geq 0$$

$$x(t) - y(t) \geq 0$$

for all $t \in [0,T]$. Hence both $x + y$ and $x - y$ satisfy (16). Since y is not identically zero, it follows that x is not an extreme point. $\square$

2.3. Definition of a basic feasible solution

With the above characterization of right analytic extreme points, we can now make precise the notion of basic feasible solutions in continuous time. This will be fundamental in laying the ground work that will be used in the remaining chapters to develop a continuous time simplex method.

From now on we shall concern ourselves only with extreme points having finitely many breakpoints.

(17) Definition. Let $x:[0,T] \to R^n$ satisfy

$$Bx(t) + K\int_0^t x(s)ds = b(t)\ , \qquad x(t) \geq 0,\ t \in [0,T]$$

where $B, K \in R^{m\times n}$. Then x will be called a basic feasible solution if there are finitely many points

$$0 = t_0 < t_1 < \cdots < t_k = T$$

and for each $j = 1,\ldots,k$, a partition (α_j,β_j) of the variables such that

(i) $|\beta_j| = m$

(ii) $x_{\alpha_j}(t) = 0,\ t \in (t_{j-1},t_j)$

(iii) $\det(\mu B_{\cdot\beta_j} + K_{\cdot\beta_j}) \neq 0$ for some μ.

If in addition $x_{\beta_j} > 0$ on the open interval (t_{j-1},t_j) for each j, and $x_r(t_j^-) > 0$, $x_r(t_j^+) > 0$ for $r \in \beta_j \cap \beta_{j+1}$, then x will be called non-degenerate. On the interval $[t_{j-1},t_j)$, x_{β_j} and x_{α_j} will be called basic and nonbasic variables respectively.

Note that if $b(\cdot)$ is analytic then by the result of Proposition C.3(18) any basic feasible solution x will, for each interval (t_{j-1},t_j), agree with some function $y^j(\cdot)$ that is analytic

on a neighborhood of $[t_{j-1}, t_j]$. In particular this implies that both the left and right limits, $x^{(i)}(t_j^-)$ and $x^{(i)}(t_j^+)$ exist for all $j = 1,\ldots,k$ and all $i = 0,1,2,\ldots$.

CHAPTER 3

THE SIMPLEX METHOD IN R^n AND CONTINUOUS TIME VERSUS DISCRETE TIME

The purpose of this chapter is to motivate the approach we shall take in developing the theory of a simplex method in continuous time. In Chapter 2 we saw that it was possible to characterize all right analytic extreme points as "basic feasible solutions" in a manner very much akin to that of linear programming. Here we show that the naive generalization of the simplex method to continuous time is inadequate for this class of problems and that additional new techniques are required.

3.1. The Simplex Method in R^n

We first briefly describe a single iteration in the simplex method. While this is well-known [6], we present it to illustrate and motivate the approach in continuous time.

The problem is

$$\text{(1)} \qquad \begin{array}{l} \text{minimize: } cx \\ \text{subject to: } Ax = b,\ x \geq 0,\ x \in R^n \end{array}$$

Let x be a given feasible solution. Then for any λ

$$cx = cx + \lambda(b - Ax) .$$

Upon rearrangement this becomes

$$(2) \qquad cx = (c - \lambda A)x + \lambda b.^{1}$$

To obtain an algorithm, the following condition is required.

(3) <u>Complementary slackness condition</u>.

$$x_i > 0 \Rightarrow c_i - \lambda A_{\cdot i} = 0 , \qquad i = 1,\ldots,n$$

The following simple but fundamental lemma follows immediately by inspection.

(4) <u>Lemma</u>. Let x and λ be given as above and suppose that the complementary slackness condition (3) holds. Then if $c_s - \lambda A_{\cdot s} < 0$, and x_s (currently at zero) can be increased so that <u>only</u> the positive components of x need be adjusted to maintain the relations $Ax = b$, $x \geq 0$, the new solution obtained in this way will have a lower objective value. $\square$

The simplex method uses this lemma in the following way.

Begin with any extreme point x, and let $\beta = \{i:x_i > 0\}$. For this discussion we assume "nondegeneracy," i.e. the matrix $A_{\cdot\beta}$ is square and nonsingular. Proceed now as follows:

<u>Step 1</u>. <u>Solve for the prices</u>. Require complementary slackness, i.e. $c_\beta - \lambda A_{\cdot\beta} = 0$, and solve uniquely for λ.

---

[1] $c - \lambda A$ is called the reduced cost of x.

Step 2. "Pricing": Select ℓ such that $c_\ell - \lambda a_\ell < 0$. If no such ℓ exists we are optimal--stop.

Step 3. Represent the incoming activity: Solve the system $A_{\cdot\beta} y_\beta = a_\ell$.

Step 4. "Minimum Ratio": Increasing x_ℓ to some level θ yields a new x_β given by

$$x'_\beta = x_\beta - y_\beta \theta$$

To obtain the maximum decrease in the objective, increase θ as far as possible, but maintaining $x'_\beta = x_\beta - y_\beta\theta \geq 0$. Under nondegeneracy precisely one component of x'_β, say x'_r, will be zero. r is determined by

$$r = \operatorname{argmin}\left\{\frac{x_i}{y_i} : y_i > 0\right\}$$

and the maximum θ is given by

$$\theta = \frac{x_r}{y_r}$$

where we set $y_{\sim\beta} = 0$. Define the new basis to be $\beta' = \beta \cup \{\ell\} \sim \{r\}$, set $x'_\ell = \theta$, and proceed to Step 1. □

3.2. Continuous Time Versus Discrete Time

A priori there appears little reason why, without any further qualification, we cannot apply the above simplex steps to the continuous time problem. The point to note, however, is that if we are to work with solutions having activities basic over intervals of time, then

in order to get a decrease in the objective value by a change of basis, this change will have to occur over an interval(s) of time. The discrete time analog of the above process would correspond to a simultaneous exchange of many activities outside the basis with those in the basis. This differs from the simplex method where we do these exchanges one at a time.

In this case, however, Lemma (4) still applies: By a simultaneous increase of several activities having negative reduced costs, we can still obtain a strict decrease provided we only adjust existing positive (basic) activities to maintain the constraints. What goes 'wrong' in continuous time is that without some basic variables becoming negative, this can seldom be done. In order to make a change of basis over an interval of time, we usually in addition have to adjust nonbasic activities, which may well have positive reduced costs. However, once we do this there is no immediate reason why the intended change of basis should yield an improvement.

Let us illustrate with an example.

(5) Example.

$$\text{minimize:} \quad \int_0^2 \{(t-2)x_1 + 2x_2 + tx_3\}dt$$

$$\text{subject to:} \quad \lambda_1: \; x_1(t) - \int_0^t x_2(s)ds = 1-t$$

$$\lambda_2: \; x_2(t) + x_3(t) = 2$$

$$x_i(t) \geq 0, \quad t \in [0,2].$$

We shall first show that what happens in continuous time and then interpret this carefully in discrete time.

Begin with the following basic solution:

$$
(6)\qquad
\begin{aligned}
&[0,1):\ x_1(t) = 1 - t,\ x_2(t) = 0,\ x_3(t) = 2 \\
&[1,2]:\ x_1(t) = 0\ ,\qquad x_2(t) = x_3(t) = 1
\end{aligned}
$$

By requiring complementary slackness over these two intervals, we can uniquely solve[1] for the prices $\lambda(\cdot)$, and obtain:

$$
\begin{aligned}
(1,2]:\ &\lambda_1(t) = -1 \\
&\lambda_2(t) = t \\
&\bar{c}_1(t) = t-1 > 0
\end{aligned}
$$

$$
\begin{aligned}
[0,1]:\ &\lambda_1(t) = t-2 \\
&\lambda_2(t) = t \\
&\bar{c}_2(t) = -\tfrac{1}{2}(t-1)^2 \leq 0
\end{aligned}
$$

where $\bar{c}_i(\cdot)$ is the reduced cost of the nonbasic variable x_i on the appropriate interval.

Examination of the reduced costs shows that we should increase x_2 in the first interval. Suppose we do this by exchanging x_2 and x_3 on the interval $[0,\epsilon)$, $\epsilon > 0$. This will yield the following solution on $[0,1)$:

[1]This procedure will be well defined later on.

$$[0,\epsilon): \; x_1(t) = 1 + t, \qquad x_2(t) = 2, \; x_3(t) = 0$$

$$[\epsilon,1): \; x_1(t) = 1 + 2\epsilon - t, \; x_2(t) = 0, \; x_3(t) = 2$$

To obtain the new solution on $[1,2]$, we would like to maintain basic the same variables as in the original solution, i.e. x_2 and x_3. Thus, keeping $x_1 = 0$ over $[1,2]$, we require x_2 and x_3 to satisfy

$$\left.\begin{array}{l} -\int_1^t x_2(s)ds = 1 - t + 2\epsilon \\ x_2(t) + x_3(t) = 2. \end{array}\right. \qquad t \in [1,2]$$

However, the only solution to these equations is

$$\left.\begin{array}{l} x_2(t) = 1 - 2\epsilon\delta(t-1) \\ x_3(t) = 1 + 2\epsilon\delta(t-1) \;, \end{array}\right. \qquad t \in [1,2]$$

where $\delta(t)$ is the delta functional.[1]

We observe immediately that $x_2(t) = -\infty$ at $t = 1$, and thus x_2 there fails to satisfy the nonnegativity restriction. Hence it is not possible to adjust only the basic variables and remain feasible.

If we did not require $x_1 = 0$ over $[1,2]$, and instead held it fixed at $x_1(t) = 2\epsilon$, its value immediately prior to $t = 1$, we would obtain the solution

$$\left.\begin{array}{l} x_1(t) = 2\epsilon \\ x_2(t) = 1 \\ x_3(t) = 1 \;. \end{array}\right. \qquad t \in [1,2]$$

[1] δ may be thought of as a "function" that is zero everywhere except at the origin where it is so large that it integrates to 1. See Appendix C.

However, this solution is not a basic solution and moreover has an objective value that is greater than that of the original solution. This is because the reduced cost of x_1 over $[1,2]$ is positive, and in this case keeping x_1 at 2ϵ over $[1,2]$ yields a positive contribution to the objective that dominates the decrease obtained over $[0,\epsilon)$. See Figure 1 below.

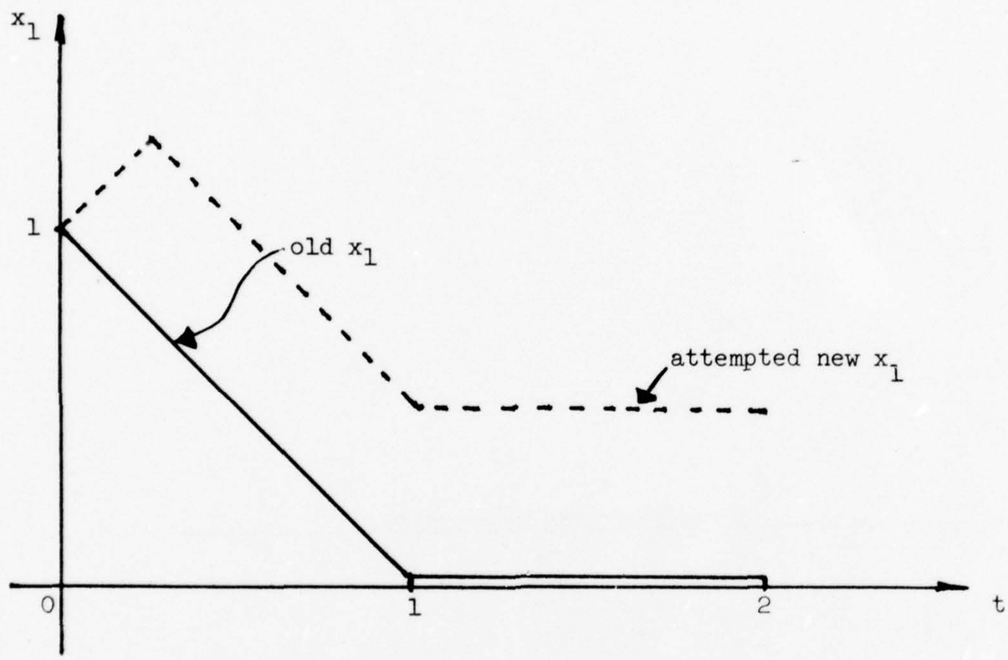


FIGURE 1

The approach we shall pursue will be to allow the breakpoint at $t = 1$ to adjust so that x_1 can be allowed to reach zero continuously and then be kept at zero level from then onwards. This will yield the following solution:

$$[0,\epsilon): \quad x_1(t) = 1 + t \qquad x_2(t) = 2,\ x_3(t) = 0\ ,$$

$$[\epsilon,\ 1+2\epsilon): \quad x_1(t) = 1 + 2\epsilon - t, \quad x_2(t) = 0,\ x_3(t) = 2\ ,$$

$$[1+2\epsilon,\ 2): \quad x_1(t) = 0, \qquad x_2(t) = 1,\ x_3(t) = 1\ .$$

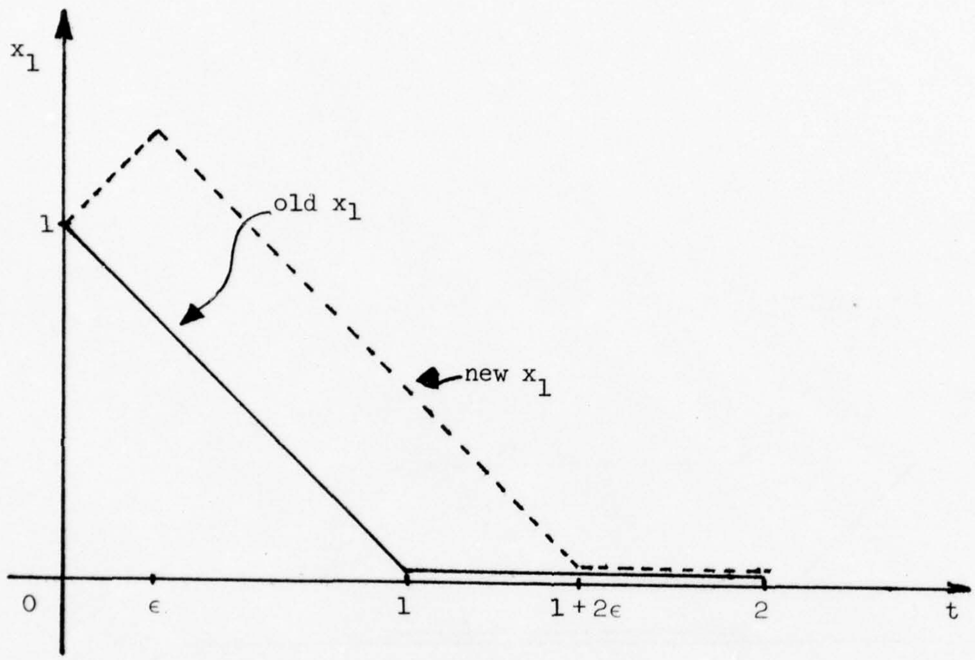


FIGURE 2

Here we have preserved on $t \geq \epsilon$ the same sequence of basic variables as before and have only altered the timing of the change over from $\{x_1, x_3\}$ to $\{x_2, x_3\}$. Thus we end up with a new basic feasible solution. In addition, even though x_1 is now positive over the interval $[1,\ 1 + 2\epsilon)$ where it has a positive reduced cost, it turns out upon recalculation of the objective value that an overall strict decrease is achieved for ϵ sufficiently small. Indeed,

setting $\epsilon = 1/3$ yields the optimal solution (easily verified by recomputing the new reduced costs).

If, for a given ϵ sufficiently small, we consider this to be a single iteration, then it is certainly not a case to which Lemma (4) applies. However, as the analysis of the discrete case will show, we can think of this as an infinite number of basis changes, switching back and forth between $t = 0^+$ and $t = 1^+$, and accompanied by repricing an infinite number of times.

To discretize the problem, we divide the time interval $[0,2]$ into N intervals of equal size, and approximate the integral using the leftmost function values. Let $\delta = 2/N$.[1] Then the discrete formulation becomes

$$\text{minimize:} \quad \sum_{k=0}^{N-1} \{(k\delta - 2)\, x_1(k\delta) + 2x_2(k\delta) + k\delta x_3(k\delta)\}\delta$$

subject to:

$$\lambda_1(r\delta): \quad x_1(r\delta) = \sum_{k=0}^{r-1} x_2(k\delta)\delta + (1 - r\delta)$$

$$\lambda_2(r\delta): \quad x_2(r\delta) + x_3(r\delta) = 2$$

$$x_i(r\delta) \geq 0, \qquad r = 0,1,\ldots,N$$

The detached coefficient tableau is given in Table 1.

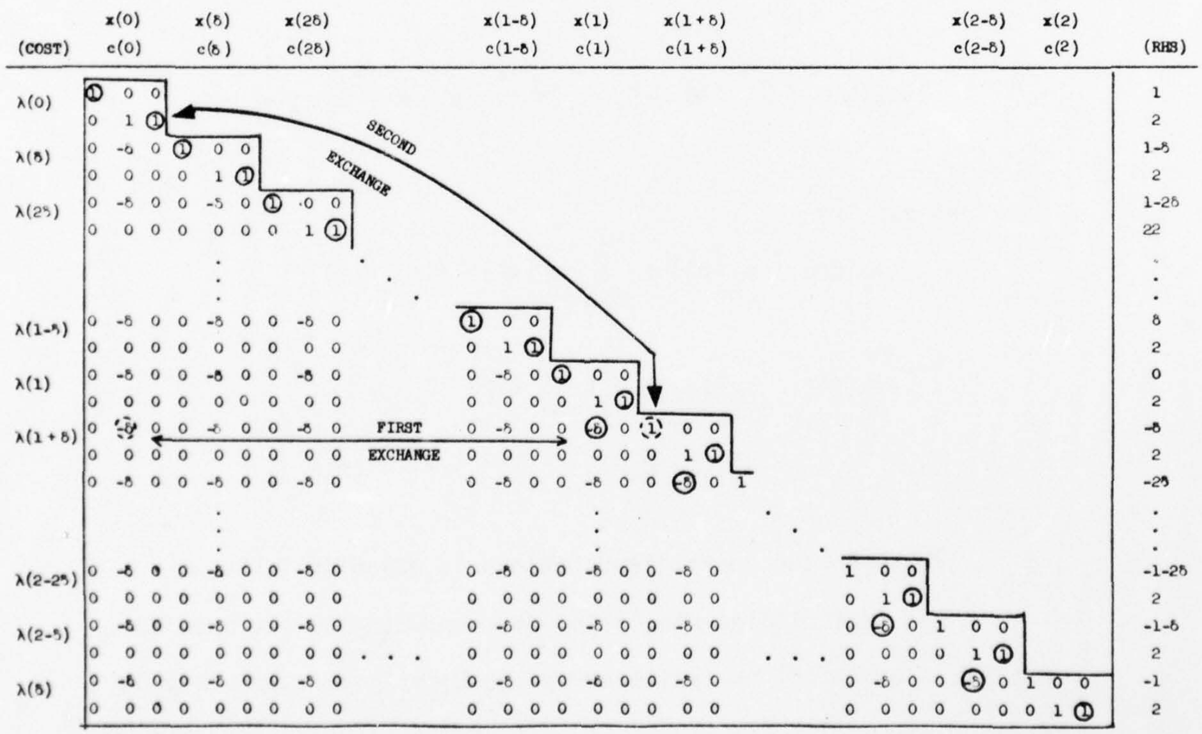
The initial discrete basis, corresponding to the initial continuous basis (6) is indicated by the heavily circled elements in the tableau. These are the pivotal elements. Note that for

[1] Not to be confused with the δ functional.

$t < 1$, x_1 and x_3 are basic; at $t = 1$, all of x_1, x_2, x_3 are basic; for $1 < t < 2$, x_2 and x_3 are basic, and at $t = 2$, only x_3 is basic. Thus in the discrete case, we have a surplus basic variable at $t = 1$ and are one short at $t = 2$.

We now solve backwards in time for the prices. As in the continuous case it turns out that $x_2(0)$ has a negative reduced cost. Increasing $x_2(0)$ has the effect of dropping $x_2(1)$ from the basis. This exchange is indicated in Table 1 where the element $-\delta$ with the dotted circle is not the new pivotal element for $x_2(0)$. Note that the surplus at $t = 1$ has now shifted to $t = 0$.

Table 1



$c(t) = [t-2,\ 2,\ t]$

The next step is to solve again for the prices. However in this case we cannot sequentially solve backwards for $\lambda(\cdot)$, since the value of $\lambda_1(1+\delta)$ can only be determined once the value of $\lambda_2(0)$ has been found. Solving yields $\lambda_1(1+\delta) = \frac{3}{2\delta} - \frac{1}{2}$, so that as $\delta \to 0$, $\lambda_1(1+\delta) \to \infty$. The reduced costs of the nonbasic variables are all of order unity except for that of $x_1(1+\delta)$ which has value $\delta - \frac{3}{2} - \frac{3}{2\delta} \ll 0$. Hence we now introduce $x_1(1+\delta)$ into the basis. This has the effect of dropping $x_3(0)$, and so transfers the basis surplus from $t = 0$ to $t = 1 + \delta$. See Table 1.

Below we sketch the effect of these two iterations on $x_1(\cdot)$.

Clearly, the simplex method in discrete time is behaving in the same way as the proposed step in the continuous case for small ϵ.

The remainder of this thesis will be devoted to making rigorous the above principle in continuous time, viz.: if an activity prices out favorably, then without further pricing, it can be increased over a sufficiently small interval to yield a decrease in the objective value.

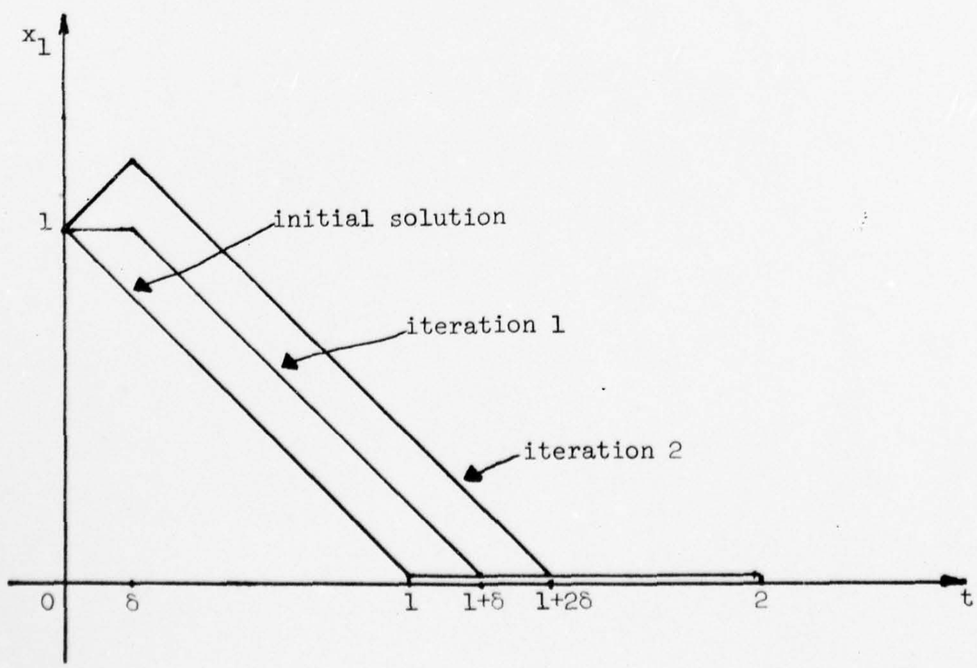


FIGURE 3

CHAPTER 4

ON A CONTINUOUS TIME SIMPLEX METHOD

In this chapter we develop a theory for a continuous time simplex method. First we consider the optimality conditions and the notion of a reduced cost in continuous time. Then we show how to move from one basic feasible solution to a nearby one with a lower objective value.

4.1. Optimality Conditions and Weak Duality

As in the simplex method, or most other constrained optimization problems, the first step is to modify the cost functional by adding to it certain linear combinations of the constraints.

Recall that our problem is

$$\text{minimize:} \quad \int_0^T c(t)\, x(t)dt$$

$$\text{subject to:} \quad Bx(t) + K\int_0^t x(s)ds = b(t) \tag{1}$$

$$x(t) \geq 0, \quad t \in [0,T]$$

where we assume that $b(\cdot)$ and $c(\cdot)$ are analytic, and $B, K \in R^{m \times n}$.

The following Lemma is easily obtained.

(2) Lemma. Let $x(\cdot)$ satisfy (1), and let $0 = t_0 < t_1 < \cdots < t_k = T$ be any points such that the derivatives $x^{(j)}(t_i -)$ exist for all $i = 1,\ldots,k$ and $0 \leq j \leq \ell$.[1] Let $\lambda^*:[0,T] \to R^m$ be any function[2] and ν^{ji}, $j = 0,1,\ldots,\ell$, $i = 1,\ldots,k$ be any vectors in R^m. Then

$$(3) \quad \int_0^T c(t)\, x(t)dt$$

$$= \int_0^T \lambda^*(t)\, b(t)dt + \sum_{i=1}^{k} \sum_{j=0}^{\ell} \nu^{ji}\, b^{(j)}(t_i)$$

$$+ \int_0^T \bar{c}(t)\, x(t)dt - \sum_{i=1}^{k} \sum_{j=0}^{\ell} \bar{d}^{ji}\, x^{(j)}(t_i^-)$$

where $\bar{c}(\cdot)$ and $\{\bar{d}^{ji}\}$ are defined by

$$(4) \qquad \bar{c}(t) = c(t) - \lambda^*(t)B - \int_t^T \lambda^*(s)ds\, K - \sum_{i=1}^{k} \nu^{0i}\, K\, 1_+(t_i - t)$$ [3]

$$(5) \qquad \bar{d}^{ji} = \nu^{ji}B + \nu^{j+1,i}K$$

and we define $\nu^{ji} = 0$ for $j > \ell$.

Proof. By assumption we may differentiate the constraints (1) and evaluate all quantities at $t = t_i^-$. Thus

---

[1] Some k, ℓ.

[2] We assume both x and λ^* are integrable.

[3] $1_+(t) = 0$ if $t < 0$
$= 1$ if $t \geq 0$.

$$b^{(j)}(t_i) - Bx^{(j)}(t_i^-) - Kx^{(j-1)}(t_i^-) = 0$$

$$0 \leq j \leq \ell, \quad 1 \leq i \leq k$$

where we define

$$x^{(-1)}(t) = \int_0^t x(s)ds \ .$$

Now left multiply the (i,j)th relation by ν^{ji} and add it to the cost functional. Also, left multiply (1) by $\lambda^*(t)$, integrate over $[0,T]$, and add that result to the cost functional. This yields

$$\int_0^T c(t)\ x(t)dt$$

$$= \int_0^T c(t)x(t)dt + \int_0^T \lambda^*(t)\{b(t) - Bx(t) - K\int_0^t x(s)ds\}dt$$

$$+ \sum_{k=1}^{k} \sum_{j=0}^{\ell} \nu^{ji}\{b^{(j)}(t_i) - Bx^{(j)}(t_i^-) - Kx^{(j-1)}(t_i^-)\}$$

Interchanging the order of integration and regrouping terms yields the result. $\square$

We remark that Lagrangian constructions of this type considered by previous authors in this area (see Chapter 1) have been concerned only with the case $\nu^{ji} = 0$. For our purposes, however, allowing them to be nonzero is essential. We shall later see that the ν^{ji} can be interpreted as the coefficients of δ's and their higher order derivatives, $\delta^{(j)}$.[1]

[1]See Appendix C for definition of $\delta^{(j)}$.

Observing that $x(t) \geq 0$ and that the derivatives $x^{(j)}(t_i^-)$ are unrestricted in sign for $j \geq 1$, we can now use Lemma (2) to define the natural dual of problem (1), and make a statement about weak duality.

The dual problem reads as follows:

Find time points $0 = t_0 < t_1 < \cdots < t_k = T$, constant vectors $\{\nu^{ji}\} \subset R^m$, and a function $\lambda^*:[0,T] \to R^m$ so as to

$$\text{maximize:} \int_0^T \lambda^*(t)\ b(t)\ dt + \sum_{i=1}^{k} \sum_{j=0}^{\ell} \nu^{ji}\ b^{(j)}(t_i)$$

$$(6) \qquad \text{subject to:} \quad \lambda^*(t)B + \int_t^T \lambda^*(s)ds\ K + \sum_{i=1}^{k} \nu^{0i} K 1_+(t_i - t) \leq c(t)$$

$$\nu^{0i}B + \nu^{1,i}K \leq 0$$

$$\nu^{ji}B + \nu^{j+1,i}K = 0$$

$$t \in [0,T], \quad j = 0,1,\ldots,\ell, \quad i = 1,\ldots,k.$$

(7) Lemma. Weak Duality.

Let the triple $(\lambda^*, \{\nu^{ji}\}, \{t_i\})$ be any feasible solution to the dual constraints (6), and x any feasible solution to the primal constraints (1) that possesses left derivatives $x^{(j)}(t_i^-)$. Then

$$\text{(i)} \quad \int_0^T c(t)\ x(t)dt \geq \int_0^T \lambda^*(t)\ b(t)dt + \sum_{i=1}^{k} \sum_{j=0}^{\ell} \nu^{ji}\ b^{(j)}(t_i)$$

(ii) Equality holds in (i) if and only if the following conditions hold.

For $r = 1,\dots,n$ and almost all $t \in [0,T]$

(a) $x_r(t) > 0$ implies

$$\lambda^*(t)b_r + \int_t^T \lambda^*(s)ds\, k_r + \sum_{i=1}^{k} \nu^{0i} k_r 1_+(t_i - t) = c_r(t)$$

(b) $x_r(t_i^-) > 0$ implies

$$\nu^{0i} b_r + \nu^{1,i} k_r = 0$$

(iii) If equality holds in (i) then x is an optimal solution to (1) and the triple $(\lambda^*, \{\nu^{ji}\}, \{t_i\})$ is an optimal solution to (6).

<u>Proof</u>. This follows immediately from Lemma (2). □

Conditions (a) and (b) are complementary slackness conditions. In order to make them apply to any triple $(\lambda^*, \{\nu^{ji}\}, \{t_i\})$ that need not be dual feasible[1], it will be convenient in addition to make a statement about when we require the higher order terms $\nu^{ji} b_r + \nu^{j+1,i} k_r$ to be zero.

(8) <u>Definition. Complementary Slackness Condition</u>

Let $(\lambda^*, \{\nu^{ji}\}, \{t_i\})$ be any triple, and x a feasible solution to (1) that possesses left derivatives $\{x^{(j)}(t_i^-)\}$. Then

[1] The triple $(\lambda^*, \{\nu^{ji}\}, \{t_i\})$ will be said to be dual feasible if it satisfies the constraints in (6).

x and $(\lambda^*, \{\nu^{ji}\}, \{t_i\})$ will be said to satisfy the <u>complementary slackness condition</u> if (a) and (b) of Lemma (7) hold, and in addition condition (c) below holds:

(c) $x_r(\cdot)$ is not identically zero on the interval $(t_i - \delta, t_i)$ for any $\delta > 0$ implies

$$\nu^{ji} b_r + \nu^{j+1,i} k_r = 0 , \qquad 1 \leq j \leq \ell . \qquad \square$$

Note that if $(\lambda^*, \{\nu^{ji}\}, \{t_i\})$ is dual feasible and satisfies (a) and (b), then (c) is automatically satisfied.

Lemma (7) provides us with a useful means of verifying that a candidate solution x is an optimal solution. We simply pick any time points $\{t_i\}$, find (if possible) a $\lambda^*(\cdot)$ and $\{\nu^{ji}\}$ that satisfy the complementary slackness conditions, and then check to see if the triple $(\lambda^*, \{\nu^{ji}\}, \{t_i\})$ is dual feasible. If so, x is an optimal solution.

In order to exploit this fact, it will be important to use the complementary slackness condition to define the dual variables. As in the simplex method, and as we shall show here, this can be done when x is a basic feasible solution. For such an x, we use the $\{t_i\}$ given by its induced partition of $[0,T]$, and then require λ^* and $\{\nu^{ji}\}$ to satisfy

$$(9) \qquad \lambda^*(t)B._{\beta_i} + \int_t^T \lambda^*(s)ds\, K._{\beta_i} + \sum_{r=i}^{k} \nu^{0r} K._{\beta_i} = c_{\beta_i}(t)$$

$$\nu^{ji} B._{\beta_i} + \nu^{j+1,i} K._{\beta_i} = 0$$

$$0 \leq j \leq \ell, \quad t \in (t_{i-1}, t_i) \quad \text{for } i = 1,\dots,k.$$

Note that if λ^* and $\{\nu^{ji}\}$ satisfy (9) then complementary slackness holds between x and $(\lambda^*, \{\nu^{ji}\}, \{t_i\})$. However, the converse is not true in general since even a nondegenerate x may have $x_r(t_i^-) = 0$ for $r \in \beta_i$ (some r, i) which then relaxes the restriction

$$\nu^{0i} b_r + \nu^{1i} k_r = 0 \ .$$

This will become important later on in Section 4.5.

Note also that in terms of $\bar{c}(\cdot)$ and $\{\bar{d}^{ji}\}$ defined in (4) and (5), (9) reads

$$\begin{aligned} \bar{c}_{\beta_i}(t) &= 0, \qquad && t \in (t_{i-1}, t_i) \\ \bar{d}^{ji}_{\beta_i} &= 0 \ , && 0 \leq j \leq \ell \end{aligned}$$

for $i = 1,\ldots,k$.

We now show that (9) uniquely determines λ^* and $\{\nu^{ji}\}$, and interpret the $\{\nu^{ji}\}$ as the coefficients of the δ functional and its higher order derivatives, $\delta^{(j)}$.

For the following lemma we shall assume that the reader is familiar with Appendix C which contains definitions of and facts about the $\{\delta^{(j)}\}$ and more general distributions.

(10) Lemma. Let $\{t_i\}$ and $\{\beta_i\}$, $i = 1,\ldots,k$ be given, and assume that for each i there is a scalar μ such that $(\mu B._{\beta_i} + K._{\beta_i})^{-1}$ exists. Then if $\ell \geq m$,

(i) (9) has a unique solution $(\lambda^*, \{\nu^{ji}\})$, which moreover necessarily satisfies $\nu^{ji} = 0$ for $j \geq m$.

(ii) $(\lambda^*, \{\nu^{ji}\})$ is the unique solution to (9) if and only if $\lambda(\cdot)$ defined by

$$\lambda(t) = \lambda^*(t) + \sum_{i=1}^{k} \sum_{j=0}^{m} \nu^{ji}\delta^{(j)}(t_i - t)$$

is the unique distributional solution to

$$\text{(11)} \qquad \lambda(t)B._{\beta_i} + \int_t^T \lambda(s)ds\, K._{\beta_i} = c_{\beta_i}(t),\ t \in (t_{i-1}, t_i),\ i = 1,\ldots,k.$$

<u>Proof</u>. Note first that if $(\lambda^*, \{\nu^{ji}\})$ is a solution to (9) then λ defined as above is a distributional solution to (11). By Proposition C.3 (18) and the assumption that $(\mu B._{\beta_i} + K._{\beta_i})^{-1}$ exists for some μ, it follows that (11) has a unique distributional solution of the form given above for λ. Further if any such λ satisfies (11), then equating coefficients of $\delta^{(i)}$ on both sides shows that the components λ^* and $\{\nu^{ji}\}$ of λ satisfy (9). □

With the aid of Lemmas (7) and (10) our initial step in a continuous simplex method is now complete. We begin with a basic feasible solution, x, and compute the unique "prices" λ^* and $\{\nu^{ji}\}$ after requiring that (9) holds. We then evaluate the reduced costs $\bar{c}(\cdot)$ and $\{\bar{d}^{ji}\}$.

If

$$\bar{c}(t) \geq 0, \qquad t \in [0,T]$$

$$\bar{d}^{0i} \leq 0$$

$$\bar{d}^{ji} = 0$$

for $j = 1,\ldots, m$, $i = 1,\ldots,k$, then x is an optimal solution. If not, we use the fact that one of these conditions has been violated to attempt to obtain an improved solution. This will be the subject of the remainder of this chapter.

Remark on the Statement of the Dual Problem

In general the dual problem as stated in (6) will not always have an optimal solution. This is because we may in general require more than a finite number of breakpoints, and thus more general distributions than finite linear combinations of functions and the $\delta^{(i)}$. However it seems that this will only occur when the primal problem (1) has an optimal solution x with infinitely many discontinuities either in x or in $x^{(j)}$.

In the space of distributions the dual problem thus reads:

$$\text{maximize:} \quad \int_0^T b(t)\,\lambda(t)\,dt$$

$$\text{subject to:} \quad \lambda(t)B + \int_t^T \lambda(s)ds\,K \leq c(t)\ , \qquad t \in [0,T]$$

for $\lambda \in \mathcal{D}'$, where the terms with $\int$ signs have their usual distributional meanings.

Since the only signed distributions are measures, any feasible λ will have to be such that $\pi(t)$, defined by

$$\pi(t) = \lambda(t)B + \int_t^T \lambda(s)ds\, K,$$

is a measure. This corresponds precisely to the requirement

$$\nu^{ji}B + \nu^{j+1,i}K = 0, \qquad j \geq 1$$

in (6), i.e. the "distributional part" of π must vanish.

That a strong duality theorem always holds for $\lambda \in \mathcal{D}'$ is still to our knowledge an open question.

4.2. Improving the Solution When Some Reduced Cost $\bar{c}_\ell(t) < 0$

We are now in a position to consider moving from one basic feasible solution to a better one. Assume that we currently have a basic feasible solution x, and that we have computed the reduced costs $\bar{c}(\cdot)$ and $\{\bar{d}^{ij}\}$ after solving (9). The case we consider here is where $\bar{c}_\ell(t^*) < 0$ for some ℓ and some $t^* \in (t_{j-1}, t_j)$. Since $\bar{c}_\ell$ is continuous on any such interval, there will be an interval about t^* on which $\bar{c}_\ell < 0$. Without loss of generality, we shall assume that $t^* = 0$.

The following major assumptions will be made throughout the remainder of this chapter.

(I) The constraint set is bounded (in L_∞).

(II) All basic solutions encountered are nondegenerate, as defined in Section 2.3.

We shall now attempt to introduce x_ℓ into the basic set over $[0,\epsilon)$, some sufficiently small $\epsilon > 0$, in such a way that a decrease in the objective and a new basic feasible solution are achieved. This will be done in two stages. The first will be to make a local basis change on $[0,\epsilon)$ without regard to what happens beyond $t = \epsilon$. The second stage will consist of adjusting the solution on $t \geq \epsilon$ (if necessary and/or possible) so that our new solution satisfies the constraints.

It will be important to remember throughout that all constructions depend on ϵ. For clarity of exposition, this dependence will not be written explicitly.

4.2.1. Making a local change over $[0,\epsilon)$.

On the interval $[0,t_1)$, the first constant basis interval, the current solution x satisfies

$$(12) \qquad B_{\cdot\beta_1} x_{\beta_1}(t) + K_{\cdot\beta_1} \int_0^t x_{\beta_1}(s)ds = b(t)$$

$$x_{\alpha_1}(t) = 0$$

If we are to increase x_ℓ, then the new solution, to be denoted by x', will have to satisfy

(13) $$B._{\beta_1} x'_{\beta_1}(t) + K._{\beta_1} \int_0^t x'_{\beta_1}(s)ds$$

$$= b(t) - b_\ell \theta(t) - k_\ell \int_0^t \theta(s)ds$$

where θ denotes the level of x'_ℓ.

Since x is a basic solution, there is a μ such that $(\mu B._{\beta_1} + K._{\beta_1})^{-1}$ exists. Thus to solve for x'_{β_1} in terms of θ, we can apply the result of Proposition C.3(18). However the expressions become cumbersome; instead we rework the solution directly using the techniques of Appendix C.

Taking Laplace transforms on both sides of (12) and (13), and multiplying both sides by $(B._{\beta_1} + \frac{1}{s} K._{\beta_1})^{-1}$ yields

$$\hat{x}_{\beta_1}(s) = (B._{\beta_1} + \frac{1}{s} K._{\beta_1})^{-1} \hat{b}(s)$$

and

$$\hat{x}'_{\beta_1}(s) = (B._{\beta_1} + \frac{1}{s} K._{\beta_1})^{-1} \hat{b}(s) - (B._{\beta_1} + \frac{1}{s} K._{\beta_1})^{-1} (b_\ell + \frac{1}{s} k_\ell) \hat{\theta}(s).$$

Thus

(14) $$\hat{x}'_{\beta_1}(s) = \hat{x}_{\beta_1}(s) - \hat{h}(s)\, \hat{\theta}(s)$$

where

$$\hat{h}(s) = (B._{\beta_1} + \frac{1}{s} K._{\beta_1})^{-1} (b_\ell + \frac{1}{s} k_\ell) .$$

Now $\hat{h}(s)$ is an m-vector of rational function in s, each having the same denominator. Using a partial fraction expansion of $\hat{h}(s)$, we obtain, as in C.3(14),

(15) $$\hat{h}(s) = \sum_{k=0}^{q} u_k s^k + \sum_{k=1}^{p} \sum_{r=0}^{m_k-1} v_{kr} \frac{1}{(s-\xi_k)^{r+1}}$$

for $\{u_k\} \subset R^m$, $\{v_{kr}\} \subset \mathbb{C}^m$, $\{\xi_k\} \subset \mathbb{C}$.

The following Lemma allows us to assume that there is some $u_k \neq 0$, $0 \leq k \leq m$.

(16). Lemma. Suppose there is a subset γ of the rows of B and K such that $K_{\gamma\cdot} = 0$ and every column of $B_{\gamma\cdot}$ has a nonzero element. Then $u_0 \neq 0$ where $\{u_k\}$ is given by (15).

Proof. From (14), $\hat{h}$ satisfies

$$(B_{\cdot\beta_1} + \frac{1}{s} K_{\cdot\beta_1}) \hat{h}(s) = (b_\ell + \frac{1}{s} k_\ell) .$$

Since $K_{\cdot\gamma} = 0$, the equations γ read

$$B_{\gamma\beta_1} \hat{h}(s) = b_{\gamma\ell} .$$

By assumption, $b_{\gamma\ell} \neq 0$. Substituting (15) into this relation and equating coefficients on both sides yields $B_{\gamma\beta_1} u_0 \neq 0$, i.e. $u_0 \neq 0$, as required. □

In our case we have made the assumption (I) that the constraint set is bounded. Thus, if B and K do not already satisfy the condition of Lemma (16) we can either add sufficiently large simple upper bounds on the variables, or add a single row

$$ex(t) + v(t) = M$$

where $v(t) \geq 0$ is a slack variable, e is a row vector of ones, and M is a large constant. In either case we obtain the required form for B and K. Accordingly, in (15) we shall let q be the largest integer such that $u_q \neq 0$, where now $0 \leq q \leq m$.

Following Section C.3 we can take inverse Laplace transforms and use (14) and (15) to obtain

$$(17) \qquad x'_{\beta_1}(t) = x_{\beta_1}(t) - \sum_{r=0}^{q-1} \left(\sum_{k=r+1}^{q} u_k \theta^{(k-r-1)}(0^+) \right) \delta^{(r)}(t)$$

$$- \sum_{k=0}^{q} u_k \theta^{(k)}(t) - \int_0^t \psi(t-s)\, \theta(s)ds, \quad t \geq 0$$

where

$$(18) \qquad \psi(t) = \sum_{k=1}^{p} \sum_{r=0}^{m_k-1} v_{kr} \frac{1}{r!} t^r e^{\xi_k t}, \qquad t \geq 0 .$$

Expression (17) tells us precisely how to go about choosing the appropriate θ over the interval $[0,\epsilon)$. The first point to note is that the boundedness assumption (I) implies that all solutions are functions. Accordingly, our choice of θ will have to be such that the coefficients of $\{\delta^{(k)}\}$ in (17) vanish. Thus at $t = 0^+$, θ must satisfy

$$\sum_{k=r+1}^{q} u_k \theta^{(k-r-1)}(0^+) = 0 , \qquad r = 0,1,\ldots,q-1$$

i.e.

$$u_q\theta^{(0)}(0^+) = 0$$

$$u_{q-1}\theta^{(0)}(0^+) + u_q\theta^{(1)}(0^+) = 0$$

$$\vdots$$

$$u_1\theta^{(0)}(0^+) + \cdots + u_q\theta^{(q-1)}(0^+) = 0 .$$

Since $u_q \neq 0$, this implies that

(19) $$\theta(0^+) = \theta^{(1)}(0^+) = \cdots = \theta^{(q-1)}(0^+) = 0 .$$

Clearly, if $q = 0$, i.e. the terms in $\delta^{(k)}$ in (17) are absent, then these restrictions on $\{\theta^{(k)}(0^+)\}$ do not apply. As mentioned in Appendix C, a sufficient condition for this to occur is that $B_{\cdot\beta_1}^{-1}$ exists.

Next we consider what happens at $t = \epsilon$. Since we are making θ basic only over $[0,\epsilon)$, we must have $\theta(t)$ identically zero for t in an interval to the right of ϵ. This implies that $\theta^{(k)}(t) = 0$ for all such $t \geq \epsilon$. However to meet this condition without having to introduce δ's at $t = \epsilon$, the term

$$\sum_{k=0}^{q} u_k\theta^{(k)}(t)$$

in (17) implies that $\theta^{(k)}$, $k = 0,1,\ldots,q-1$, must be continuous at $t = \epsilon$, i.e. we require

$$(20) \qquad \theta(\epsilon^-) = \theta^{(1)}(\epsilon^-) = \cdots = \theta^{(q-1)}(\epsilon^-) = 0 \ .$$

Once again, if $q = 0$ then these restrictions do not apply.

Since $\bar{c}_\ell < 0$ on $[0,\epsilon)$, ϵ sufficiently small, we wish to have θ in some sense as large as possible over $[0,\epsilon)$. Clearly any feasible θ that is positive over $(0,\epsilon)$ and satisfies the restrictions (19) and (20) will yield a local decrease in the objective. However, we need first of all to choose θ such that the resulting solution is again a basic solution. We shall do this by thinking of $\theta^{(q)}$ as a bang-bang control variable that will alternate over successive subintervals of $[0,\epsilon)$ between positive and negative levels that are as large as possible.

In order, secondly, to satisfy the restrictions (20) we shall require at least q switching times for change of sign of $\theta^{(q)}$ in the interval $(0,\epsilon)$. We shall here make use of exactly q such time points. Denote these by $\tau_1\epsilon, \tau_2\epsilon, \ldots, \tau_q\epsilon$, $0 < \tau_1 < \cdots < \tau_q < 1$. We now proceed as follows. At $t = 0$, (17) reads

$$x'_{\beta_1}(0) = x_{\beta_1}(0) - u_q\theta^{(q)}(0)$$

assuming that (19) holds. We increase $\theta^{(q)}(0)$ as far as possible, maintaining

$$x_{\beta_1}(0) - u_q\theta^{(q)}(0) \geq 0 \ .$$

Under nondegeneracy exactly one component of $x'_{\beta_1}(0)$, say $x'_{r_1}(0)$, will become zero. This is exactly the same as the minimum ratio test in linear programming. r_1 is given by

(21) $$r_1 = \text{argmin}\left\{\frac{x_i(0)}{y_i} : y_i > 0\right\}$$

where $y_{\beta_1} = u_q$, $y_{\alpha_1} = 0$. Note that the boundedness condition implies that u_q has a positive component since otherwise we could increase $\theta^{(q)}$, and hence $x'_{\beta_1}(0)$, indefinitely.

Let $\sigma_1 = \beta_1 \cup \{\ell\} \sim \{r_1\}$ be the new basis at $t = 0$. We now wish to maintain this basis change over the interval $[0, \tau_1 \epsilon)$, assuming for the moment that τ_1 is given. To do this we need to check the following: (i) there is a μ such that $(\mu B._{\sigma_1} + K._{\sigma_1})^{-1}$ exists, and (ii) $x'_{\sigma_1}(t) \geq 0$ for $t \in [0, \tau_1 \epsilon)$.

(22) <u>Lemma</u>. If r_1 is determined by (21) then there is a μ such that $(\mu B._{\sigma_1} + K._{\sigma_1})^{-1}$ exists.

<u>Proof</u>. We shall use Laplace transforms. Let p be such that, from (21),

(23) $$u_{qp} = y_{r_1} > 0$$

where u_{qp} is the pth component of u_q. Write

(24) $$(B._{\sigma_1} + \frac{1}{s} K._{\sigma_1}) = (B._{\beta_1} + \frac{1}{s} K._{\beta_1}) + [(b_\ell + \frac{1}{s} k_\ell) - (b_{r_1} + \frac{1}{s} k_{r_1})]e_p^T .$$

This holds by definition of σ_1. By factoring out the matrix $(B._{\beta_1} + \frac{1}{s} K._{\beta_1})$ on the right hand side, and noting that $(b_{r_1} + \frac{1}{s} k_{r_1})$ is its pth column, we get

$$(25) \qquad (B_{\cdot\sigma_1} + \frac{1}{s} K_{\cdot\sigma_1}) = (B_{\cdot\beta_1} + \frac{1}{s} K_{\cdot\beta_1})[I + (\hat{h}(s) - e_p)e_p^T].$$

Now the matrix $I + (\hat{h}(s) - e_p)e_p^T$ is the identity with its pth column replaced by $\hat{h}(s)$. Thus it will be nonsingular iff $\hat{h}_p(s) \neq 0$. From (15) we see that

$$\hat{h}_p(s) = u_{qp}s^q + \text{lower order terms} .$$

Since, by (23), $u_{qp} \neq 0$, we obtain that $\hat{h}_p(s) \neq 0$ for s sufficiently large. This implies that for s sufficiently large, the left hand side of (25) is the product of two nonsingular matrices, and hence is itself nonsingular. □

To obtain $x'_{\sigma_1}(t)$ we can either solve the equation

$$(26) \qquad B_{\cdot\sigma_1} x_{\sigma_1}(t) + K_{\cdot\sigma_1} \int_0^t x_{\sigma_1}(s)ds = b(t)$$

from scratch, or equivalently, use (17) to compute the $\theta(t)$ that keeps $x'_{r_1}(t) = 0$ and then, again by (17), obtain the remaining components of x'_{β_1}. The latter approach yields the (q+1)th order integro-differential equation

$$(27) \qquad u_{qp}\theta^{(q)}(t) + u_{q-1,p}\theta^{(q-1)}(t) + \cdots + u_{0,p}\theta(t)$$
$$+ \int_0^t \psi_p(t-s)\ \theta(s)ds = x_{r_1}(t)$$

with boundary conditions (19).

(28) <u>Lemma</u>. For ϵ sufficiently small, $x'_{\sigma_1}(t) \geq 0$ on $[0, \tau_1\epsilon)$, any choice of $\tau_1 : 0 < \tau_1 < 1$.

<u>Proof</u>. Let $\gamma = \sigma_1 \sim \{\ell\}$. Under nondegeneracy $x'_\gamma(0) > 0$. Since x'_{σ_1} is analytic, it follows that $x'_\gamma(t) > 0$ on a neighborhood of $t = 0$. Since $\theta^{(q)}(0) > 0$ and θ satisfies the boundary conditions (19), it follows that $x'_\ell(t) = \theta(t) > 0$ for $t \in (0, \tau_1\epsilon)$. $\square$

The next step is to <u>decrease</u> $\theta^{(q)}$ as far as possible on $[\tau_1\epsilon, \tau_2\epsilon)$ while maintaining $x(t) \geq 0$ on $[\tau_1\epsilon, \tau_2\epsilon)$. This will cause the variable x_{r_1} that dropped out of the basis on $[0, \tau_1\epsilon)$ to become basic again, and another variable, x_{r_2}, to drop out of the basis. r_2 is given by

(29) $$r_2 = \operatorname{argmax} \left\{ \frac{x_i'(\tau_1\epsilon^-)}{y_i} : y_i < 0 \right\}$$

where y is as in (21). To see this note that the only discontinuity on the right hand side of (17) at $t = \tau_1\epsilon$ occurs in $\theta^{(q)}$. Thus we may write

(30) $$\begin{aligned} x'_{\beta_1}(\tau_1\epsilon^+) &= x_{\beta_1}(\tau_1\epsilon) - \sum_{k=0}^{q-1} u_k\theta^{(k)}(\tau_1\epsilon) - u_q\theta^{(q)}(\tau_1\epsilon^+) - \int_0^{\tau_1\epsilon} \psi(t-s)\theta(s)ds \\ &= x'_{\beta_1}(\tau_1\epsilon^-) - u_q[\theta^{(q)}(\tau_1\epsilon^+) - \theta^{(q)}(\tau_1\epsilon^-)] \end{aligned}$$

from which (29) follows. Note that

$$(31) \qquad x'_{r_1}(\tau_1\epsilon^+) = -u_{qp}\{\theta^{(q)}(\tau_1\epsilon^+) - \theta^{(q)}(\tau_1\epsilon^-)\} .$$

Since $u_{qp} > 0$ and $\theta^{(q)}(\tau_1\epsilon^-) > 0$, and since nondegeneracy implies $\theta^{(q)}(\tau_1\epsilon^+) < 0$ for sufficiently small ϵ, (31) verifies that $x'_{r_1}(\tau_1\epsilon^+) > 0$.

Let $\sigma_2 = \sigma_1 \cup \{r_1\} \sim \{r_2\}$. This is our new basis on $[\tau_1\epsilon, \tau_2\epsilon)$. That we indeed have a basic feasible solution on this interval for sufficiently small ϵ and τ_2 close enough to τ_1 follows from an analysis similar to that given in Lemmas (22) and (28). However, it is possible that θ may go negative if τ_2 is too large. It will be established in Corollary (4.2) below that the choice of $\{\tau_i\}$ that satisfies (32) below ensures that θ does not go negative.

We now continue in this way over the (q+1) intervals $[0,\tau_1\epsilon)$, $[\tau_1\epsilon, \tau_2\epsilon)$, ... , $[\tau_q\epsilon,\epsilon)$ alternating the sign of $\theta^{(q)}$ and obtaining (q+1) basic sets $\{\sigma_i\}$ where $\sigma_{i+1} = \sigma_i \cup \{r_i\} \sim \{r_{i+1}\}$, $i = 1,\ldots,q-1$. Each r_i is determined as in (29), and under non-degeneracy, will alternate between r_1 and r_2 for ϵ sufficiently small. Also it is easily checked that for all $\{\tau_i\}$ the x'_i remain ≥ 0 except possibly $x'_\ell = \theta$.

What remains is for us to find the points $\{\tau_i\}$ and show that for this particular choice, $\theta(t) \geq 0$ on $[0,\epsilon)$. Once this is done we shall have succeeded in making $x'_\ell = \theta$ basic over $[0,\epsilon)$, and in so doing, only have adjusted other basic variables over this interval. In doing this recall that we have not been concerned with the possibility of some basic variable becoming infeasible at some $t > \epsilon$. This will be the subject of the next section.

In principle we can compute $\{\theta^{(i)}(\epsilon^-)\}$ as functions of the parameters $\{\tau_i\}$ and then solve for the $\{\tau_i\}$ with the equations

$$
(32) \qquad
\begin{aligned}
\theta(\epsilon^- | \tau_1, \ldots, \tau_q) &= 0 \\
\theta^{(1)}(\epsilon^- | \tau_1, \ldots, \tau_q) &= 0 \\
&\vdots \\
\theta^{(q-1)}(\epsilon^- | \tau_1, \ldots, \tau_q) &= 0 \;.
\end{aligned}
$$

In practice, some iterative technique will have to be used. Below we show how a very good starting solution can be obtained.

In the following let

$$
(33) \qquad
\begin{aligned}
\rho_1 &= \min\left\{\frac{x_i(0)}{y_i} : y_i > 0\right\} \\
\rho_2 &= -\max\left\{\frac{x_i(0)}{y_i} : y_i < 0\right\}
\end{aligned}
$$

where y is as in (21). Under the boundedness condition, both ρ_1 and ρ_2 exist. Under nondegeneracy, both are positive.

(34) Theorem. For ϵ sufficiently small, equations (32) have a unique solution $0 < \tau_1 < \cdots < \tau_q < 1$. Moreover, the τ_i are given to first order by

$$
(35) \qquad \tau_i = 1 - z_i + O(\epsilon)
$$

where the z_i solve

$$
(36)\qquad
\begin{aligned}
z_1 - z_2 + \cdots + (-1)^{q+1} z_q &= \rho \\
z_1^2 - z_2^2 + \cdots + (-1)^{q+1} z_q^2 &= \rho \\
&\vdots \\
z_1^q - z_2^q + \cdots + (-1)^{q+1} z_q^q &= \rho
\end{aligned}
$$

$$1 > z_1 > \cdots > z_q > 0$$

where

$$\rho = \rho_1/(\rho_1 + \rho_2) .$$

We shall prove this theorem using the following lemmas.

(37) Lemma. Let $0 < \tau_1 < \cdots < \tau_q < 1$ be any partition of $[0,1]$. Let the bases $\{\sigma_i\}$ be determined as above. Then as $\epsilon \to 0$

$$\theta^{(q)}(t) = \rho_1 + O(\epsilon)$$

for $t \in [0, \tau_1\epsilon)$, $[\tau_2\epsilon, \tau_3\epsilon)$, ... , and

$$\theta^{(q)}(t) = -\rho_2 + O(\epsilon)$$

for $t \in [\tau_1\epsilon, \tau_2\epsilon)$, $[\tau_3\epsilon, \tau_4\epsilon)$, ...

Proof. The result on $[0, \tau_1\epsilon)$ follows directly from the analyticity of θ on this interval and the fact that $\theta^{(q)}(0) = \rho_1$. From (30),

$$\theta^{(q)}(\tau_1\epsilon^+) = \max\left\{\frac{x_i'(\tau_1\epsilon^-)}{y_i} + \theta^{(q)}(\tau_1\epsilon^-) : y_i < 0\right\} .$$

Since x' and $\theta^{(q)}$ are analytic on $[0, \tau_1\epsilon)$

$$x'(\tau_1\epsilon^-) = x'(0) + O(\epsilon)$$
$$\theta^{(q)}(\tau_1\epsilon^-) = \theta^{(q)}(0) + O(\epsilon) .$$

Also, by definition

$$x'(0) = x(0) - y\,\theta^{(q)}(0) .$$

Thus

$$\frac{x_i'(\tau_1\epsilon^-)}{y_i} + \theta^{(q)}(\tau_1\epsilon^-) = \frac{x_i(0)}{y_i} + O(\epsilon), \qquad y_i \neq 0$$

from which

$$\theta^{(q)}(\tau_1\epsilon^+) = -\rho_2 + O(\epsilon) .$$

Since $\theta^{(q)}$ is analytic on $[\tau_1\epsilon, \tau_2\epsilon)$, the result holds over this interval.

We may now continue in this way and thereby establish the lemma. □

(38) Lemma. Under the conditions of Lemma (37) θ satisfies

$$(39) \qquad \theta^{(q-j)}(t) = \rho_1 \frac{t^j}{j!} + \frac{(\rho_1 + \rho_2)}{j!} \sum_{k=1}^{i-1} (-1)^k (t-\tau_k\epsilon)^j + O(\epsilon^{j+1})$$

for $t \in [\tau_{i-1}\epsilon, \tau_i\epsilon)$, $i = 1,\ldots,q+1$, $j = 0,1,\ldots,q$, where $\tau_0 = 0$, $\tau_{q+1} = 1$.

Proof. We shall expand θ in power series that are centered at $\tau_i\epsilon$ and hold to the right of $\tau_i\epsilon$. We proceed by induction on i.

For the case i = 1, observe the boundary conditions (19), and write

$$\theta(t) = \theta^{(q)}(0)\ t^q/q! + O(t^{q+1}) .$$

Differentiating yields

$$\theta^{(q-j)}(t) = \theta^{(q)}(0)\ t^j/j! + O(t^{j+1})$$

Now $\theta^{(q)}(0) = \rho_1$ by definition. Since $t = O(\epsilon)$ for $t \in [0,\epsilon)$, the result is true.

Suppose it to be true for i. Write

$$\theta(t) = \sum_{r=0}^{q} \theta^{(r)}(\tau_1\epsilon^+)(t - \tau_i\epsilon)^r/r! + O(t^{q+1})$$

By differentiating, noting that t and $t - \tau_i\epsilon$ are $O(\epsilon)$, and using the result of Lemma (37) we obtain

$$\theta^{(q-j)}(t) = \sum_{r=0}^{j-1} \theta^{(q-j+r)}(\tau_1\epsilon^+)\ (t - \tau_i\epsilon)^r/r!$$
$$+ \gamma(t - \tau_i\epsilon)^j/j! + O(\epsilon^{j+1})$$

where

$$\gamma = \begin{cases} \rho_1 & \text{i even} \\ -\rho_2 & \text{i odd} \end{cases}$$

Using the continuity of $\theta^{(k)}$ at $t = \tau_i\epsilon$, $k = 0,1,\ldots,q-1$, and using the result for the case i, this becomes

(39) $$\theta^{(q-j)}(t)$$
$$= \sum_{r=0}^{j-1} \left\{ \rho_1(\tau_i\epsilon)^{(j-r)}/(j-r)! + \frac{(\rho_1+\rho_2)}{(j-r)!} \sum_{k=1}^{i-1} (-1)^k (\tau_i\epsilon - \tau_k\epsilon)^{j-r} + O(\epsilon^{j-r+1}) \right\} (t-\tau_i\epsilon)^r/r! + \gamma(t-\tau_i\epsilon)^j/j! + O(\epsilon^{j+1})$$

Now the following relations hold:

(i) $$\sum_{r=0}^{j-1} (\tau_i\epsilon)^{(j-r)}(t-\tau_i\epsilon)^r/(j-r)!\, r! = \frac{1}{j.}\{t^j - (t-\tau_i\epsilon)^j\}$$

(ii) $$\sum_{r=0}^{j-1}\sum_{k=1}^{i-1} (-1)^k(\tau_i\epsilon - \tau_k\epsilon)^{j-r}(t-\tau_i\epsilon)^t/(j-r)!\, r!$$
$$= \sum_{k=1}^{i-1} (-1)^k \frac{1}{j!}\{(t-\tau_k\epsilon)^j - (t-\tau_i\epsilon)^j\}$$

(iii) $$\sum_{r=0}^{j-1} O(\epsilon^{j-r+1})(t-\tau_i\epsilon)^r/r! = O(\epsilon^{j+1})$$

Substituting these into (39) yields

$$\theta^{(q-j)}(t) = \frac{\rho_1}{j!}\{t^j - (t-\tau_i\epsilon)^j\} + \frac{(\rho_1+\rho_2)}{j!}\sum_{k=1}^{i-1}(-1)^k\{(t-\tau_k\epsilon)^j - (t-\tau_i\epsilon)^j\}$$
$$+ \frac{\gamma}{j!}(t-\tau_i\epsilon)^j + O(\epsilon^{j+1})$$
$$= \rho_1\frac{t^j}{j!} + \frac{(\rho_1+\rho_2)}{j!}\sum_{k=1}^{i-1}(-1)^k(t-\tau_k\epsilon)^j$$
$$+ \frac{(\rho_1+\rho_2)(i-1)}{j!}\sum_{k=1}(-1)^{k+1}(t-\tau_i\epsilon)^j + \frac{(\gamma-\rho_1)}{j!}(t-\tau_i\epsilon)^j + O(\epsilon^{j+1})$$

By now considering the cases i even and i odd, the appropriate terms cancel and the result is established for $i+1$.

By induction, the proof is complete. $\square$

In addition to these Lemmas we shall require the results of Appendix D. These are concerned with properties of the system of equations (36).

<u>Proof of Theorem (34)</u>. Using (39), equations (32) become

$$\rho_1 \frac{\epsilon^j}{j!} + \frac{(\rho_1 + \rho_2)}{j!} \sum_{k=1}^{q} (-1)^k (\epsilon - \tau_k \epsilon)^j = O(\epsilon^{j+1})$$

for $j = 1,\ldots,q$. Thus the τ_i satisfy

$$(40) \qquad \sum_{k=1}^{q} (-1)^{k+1} (1 - \tau_k)^j = \rho + O(\epsilon), \qquad j = 1,\ldots,q$$

Now by Theorem D(2) the equations (36) have a unique solution $1 > \bar{z}_1 > \cdots > \bar{z}_q > 0$. Also by Corollary D(9) the Jacobian of the system is nonsingular for such $\{\bar{z}_i\}$. We can therefore invoke the Implicit Function Theorem [20, p. 128] and conclude that (36) has a unique solution for $O(\epsilon)$ changes in the right-hand side. Moreover this solution will not differ from $\{\bar{z}_i\}$ by more than $O(\epsilon)$. From this it follows that the solution to (40) is given by (35), and the proof is complete. $\square$

Remark. Note that since the τ_i are determined within $O(\epsilon)$, the actual breakpoints $\{\epsilon\tau_i\}$ are determined within $O(\epsilon^2)$.

We now establish that $\theta(t) \geq 0$ on $[0,\epsilon)$.

(41) **Lemma.** Let $h: R \to R$ be continuous and have a derivative that has at most finitely many simple discontinuities. If $h(t') = h(t'') = 0$ for some $t' < t''$, and $h^{(1)}(t)$ changes sign[1] k times in (t',t''), then $h(t)$ changes sign at most $k-1$ times in (t',t'').

Proof. Suppose that h changes sign r times at the points $t' < w_1 < \cdots < w_r < t''$. Then, since $h(t') = h(w_1) = \cdots = h(w_r) = h(t'') = 0$, $h^{(1)}(t)$ changes sign at least once in each interval (w_i, w_{i+1}) where $w_0 = t'$, $w_{r+1} = t''$. Thus $h^{(1)}(t)$ changes sign at least $r+1$ times in (t',t''). Hence $r+1 \leq k$, or $r \leq k-1$ as required. □

(42) **Corollary.** $\theta(t) \geq 0$ on $[0,\epsilon)$.

Proof. Since $\theta^{(i)}(0) = \theta^{(i)}(\epsilon) = 0$, $i = 0,1,\ldots,q-1$ and since $\theta^{(q)}(t)$ change sign q times in this interval, we can apply the Lemma repeatedly to obtain that θ has no change of sign in $[0,\epsilon)$. Since $\theta^{(i)}(0) = 0$ for $i = 0,1,\ldots,q-1$, and since $\theta^{(q)}(0) > 0$, the analyticity of θ ensures that $\theta(t) > 0$ in some neighborhood $(0,\delta)$. Thus $\theta(t) \geq 0$ on $[0,\epsilon)$. □

[1] By a "change of sign," we mean going from positive to negative or vice versa.

With this result, our basis change over $[0,\epsilon)$, for ϵ sufficiently small, is complete. In the event that $q = 0$, this is simply a single exchange for the duration of the interval. When $q > 0$ we require $q+1$ subintervals each having a new basis differing from its predecessor in one activity. Moreover θ is basic throughout $[0,\epsilon)$, and the bases $\{\sigma_i\}$ are all subsets of $\beta_1 \cup \{\ell\}$.

Below we illustrate the cases $q = 0,1,2$. See also Example E(1).

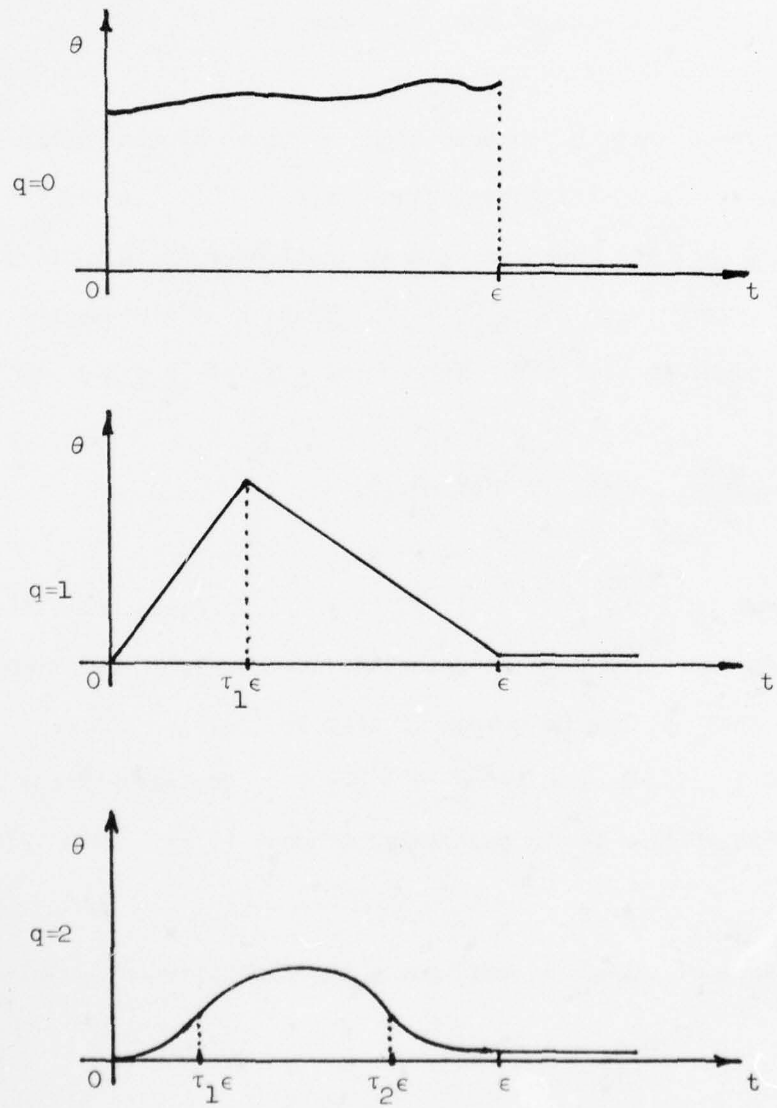


FIGURE 4

4.2.2. Adjusting the basis on $t \geq \epsilon$

So far we have obtained an x' from x that satisfies

$$Bx'(t) + K \int_0^t x'(s)ds = b(t)$$

$$x'(t) \geq 0$$

for $t \in [0,\epsilon)$. We now wish to determine the remainder of x' that satisfies these relations on $[\epsilon,T]$. There is in general no unique way of doing this, and indeed, it may not be possible at all. When no adjustment to the right of ϵ is possible, then neither is the basis change over $[0,\epsilon)$, and we shall then have to pick some other interval of perturbation. This aspect will be considered in Section 4.5.

Our approach here will be to find an x' over $[\epsilon,T]$ that is close to x in some sense, and that can be obtained in as natural a way as possible. What we shall do is try to preserve on $[\epsilon,T]$ the same sequence of basis changes as in x, and only adjust the timing of these changes. This is the approach we took in Example (5) of Section 3.2.

In addition to assumptions (I) and (II) of the previous section, we shall now assume the following:

(III) All basic solutions are such that consecutive bases β_i, β_{i+1} differ in only one element.

This assumption is a kind of nondegeneracy assumption: if β_i and β_{i+1} differ in more than one element, then we can insert bases of

zero time duration between them to achieve any two consecutive ones differing in only one element. The assumption is then that all time intervals are of positive duration.

In order to facilitate the presentation and understanding of the material in this section, we shall need to develop additional notation.

The superscript "'" will be used to denote an adjusted variable or parameter. Thus x', $\{t_i'\},\ldots$ denote the adjusted x, adjusted breakpoints,... . We remind the reader that all adjusted variables or parameters will depend on ϵ, and that this dependence will be assumed understood throughout.

ℓ_i and r_i will denote respectively the entering and leaving variables at t_i. Thus $\beta_{i+1} = \beta_i \cup \{\ell_i\} \sim \{r_i\}$.

It will be useful to work with variables $w_i(\cdot)$ that coincide with $x_{\beta_i}(\cdot)$ on $[t_{i-1},t_i)$ but that continue beyond t_i as if the breakpoints t_i, t_{i+1}, ... did not exist. Thus, given x on $[0,t_{i-1})$, w_i is defined by

$$(43)\qquad B._{\beta_i}w_i(t) + K._{\beta_i}\int_{t_{i-1}}^{t} w_i(s)ds = b(t) - K\int_0^{t_{i-1}} x(s)ds, \quad t \geq t_{i-1}$$

In addition let $\xi_i(\cdot)$ and $\eta_i(\cdot)$ be respectively those components of w_i and w_{i+1} that correspond to x_{r_i} and x_{ℓ_i}, i.e. $x_{r_i}(t) = \xi_i(t)$ on $[t_{i-1},t_i)$ and $x_{\ell_i}(t) = \eta_i(t)$ on $[t_i,t_{i+1})$.

As in the derivation of (17) we can represent x_{β_i} on $[t_i,t_{i+1})$ in terms of η_i, as follows:

$$(44) \qquad x_{\beta_i}(t) = w_i(t) - \sum_{j=0}^{m_i-1} \left(\sum_{k=j+1}^{m_i} u_k^i \, \eta_i^{(k-j-1)}(t_i^+) \right) \delta^{(j)}(t)$$

$$- \sum_{k=0}^{m_i} u_k^i \, \eta_i^{(k)}(t) - \int_{t_i}^{t} \psi_i(t-s) \, \eta_i(s) ds$$

$$t \in [t_i, t_{i+1})$$

for some $m_i \geq 0$, $\{u_k^i\}$ and $\psi_i(\cdot)$ where these are found similarly to their counterparts in (17). Here they are determined from the partial fraction expansion of

$$(B_{\cdot \beta_i} + \frac{1}{s} K_{\cdot \beta_i})^{-1} (b_{\ell_i} + \frac{1}{s} k_{\ell_i}) \quad .$$

As was done for (17), we can assume $u_{m_i}^i \neq 0$.

Of considerable importance will be the behavior of x_{r_i} and x_{ℓ_i} as they approach t_i. Let $p_i \geq 0$ be such that

$$(45) \qquad \begin{aligned} &\xi_i^{(j)}(t_i) = 0, \qquad 0 \leq j < p_i \\ &\xi_i^{(p_i)}(t_i) \neq 0 \end{aligned}$$

and $q_i \geq 0$ be such that

$$(46) \qquad \begin{aligned} &\eta_i^{(j)}(t_i^+) = 0, \qquad 0 \leq j < q_i \\ &\eta_i^{(q_i)}(t_i^+) \neq 0 \ . \end{aligned}$$

The following Lemma relates η_i and ξ_i on $t \geq t_i$.

(47) Lemma. For each $i \geq 1$

(i) $m_i \leq q_i \leq m_i + p_i$

(ii) There are $\{v_k^i\}_{k=0}^{q_i} \subset R$, $v_{q_i}^i \neq 0$, and $\varphi_i(\cdot)$ such that

$$\sum_{k=0}^{q_i} v_k^i \, \eta_i^{(k)}(t) + \int_{t_i}^{t} \varphi_i(t-s)\, \eta_i(s)ds = \xi_i^{(p_i)}(t), \qquad t \geq t_i. \tag{48}$$

Proof. Since the lhs of (44) contains no terms in $\delta^{(j)}$ the coefficients of $\delta^{(j)}$ on the rhs of (44) must vanish. Since $u_{m_i}^i \neq 0$ this implies, as in the derivation of (19), that

$$\eta^{(j)}(t_i^+) = 0\ , \qquad 0 \leq j < m_i\ .$$

By definition of q_i, this implies $q_i \geq m_i$.

Since $x_{r_i}(t) = 0$, $t \in [t_i, t_{i+1})$, (44) yields

$$0 = \xi_i(t) - \sum_{k=0}^{m_i} u_{kj}^i \eta_i^{(k)}(t) - \int_{t_i}^{t} \psi_{ij}(t-s)\, \eta_i(s)ds$$

where j is such that r_i is the jth component of β_i. Differentiating the right hand side p_i times and noting (45) yields

$$\sum_{k=0}^{m_i+p_i} u_{k-p_i,j}^i \, \eta_i^{(k)}(t) + \int_{t_i}^{t} \psi_{ij}^{(p_i)}(t-s)\, \eta_i(s)ds = \xi_i^{(p_i)}(t)$$

and

$$\sum_{k=0}^{m_i+r} u_{k-r,j}^i \, \eta_i^{(k)}(t_i^+) = 0\ , \qquad 0 \leq r < p_i \tag{49}$$

$$(50) \qquad \sum_{k=0}^{m_i+p_i} u^i_{k-p_i,j}\,\eta_i^{(k)}(t_i^+) = \xi^{(p_i)}(t_i) \neq 0$$

where

$$u^i_{-k,j} = \psi_{ij}^{(k-1)}(0)\,, \qquad 1 \leq k \leq p_i$$

From (50) there is some k such that $u^i_{k-p_i,j} \neq 0$. Let $\bar{q}_i$ be the largest such k. We show that $\bar{q}_i = q_i$. Clearly, $m_i + p_i \geq \bar{q}_i \geq q_i$, by (46). Suppose that $\bar{q}_i > q_i$. Setting $r = q_i - m_i, \ldots, p_i - 1$ in (49) and noting (46)

$$u^i_{m_i,j} = \cdots = u^i_{q_i-p_i+1,j} = 0 \quad .$$

But $u_{\bar{q}_i-p_i,j} \neq 0$ and $m_i \geq \bar{q}_i - p_i \geq q_i - p_i + 1$ yielding a contradiction. Setting $v^i_k = u^i_{k-p_i,j}$, $0 \leq k \leq q_i$ and $\varphi_i(t) = \psi_{ij}^{(p_i)}(t)$ completes the proof. $\square$

Below we sketch the cases in which p_i and q_i are one of 0, 1, or ≥ 2. Any of the pairs (p_i, q_i) is possible.

<u>Remark</u>. Note that p_i and q_i are dependent on t_i, or rather, the behavior of ξ_i and η_i at t_i. However m_i and the relation $q_i \geq m_i$ do not depend on t_i. m_i is determined solely by the coefficients $B_{\cdot\beta_i}$, $K_{\cdot\beta_i}$, b_{ℓ_i}, and k_{ℓ_i}.

We are now in a position to consider the adjustments of the t_i in order to obtain the desired x . The construction will be done inductively on i.

For $i = 0$, the adjustment has already been made: $t_0' = \epsilon$. Analogously to (43), given adjusted time points $t_0', t_1', \ldots, t_{i-1}'$,

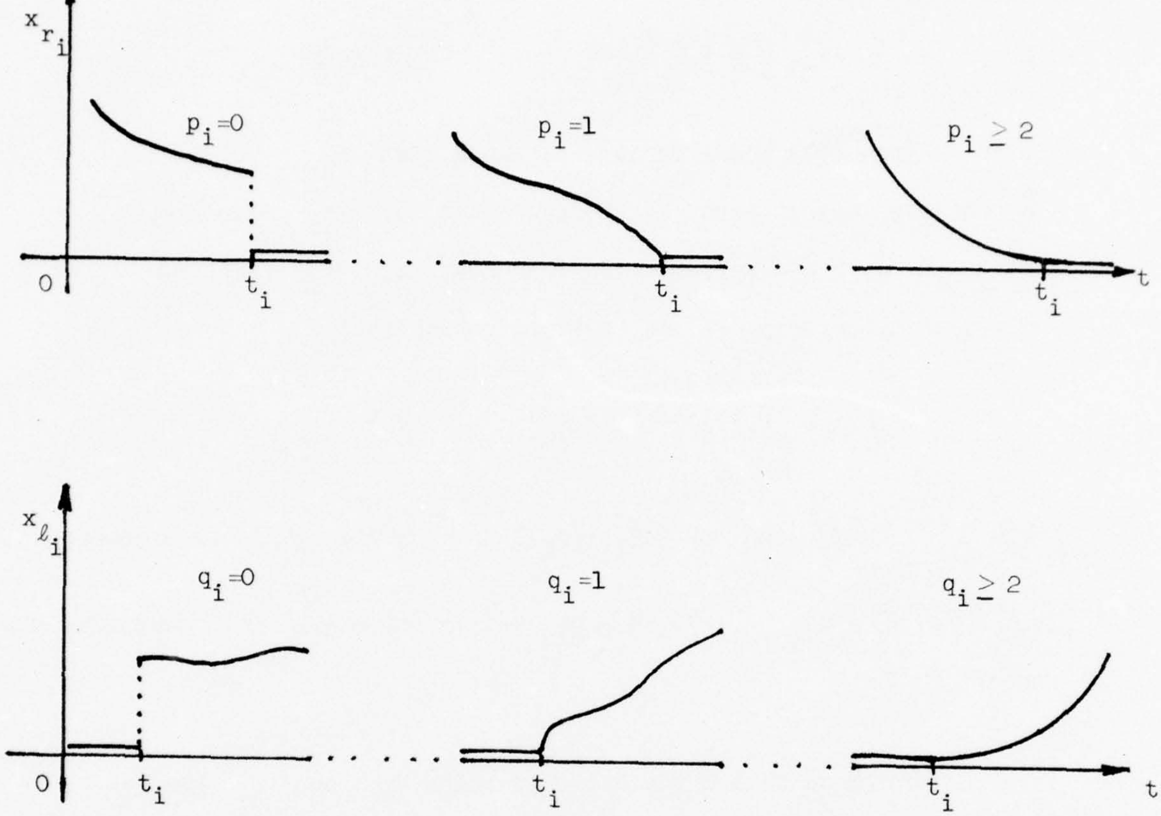


FIGURE 5

and x' defined on $[0,t'_{i-1})$ by the bases σ_j on $[\tau_{j-1}\epsilon,\tau_j\epsilon)$ $j = 1,\ldots,q+1$, and β_j on $[t'_{j-1},t'_j)$, $j = 1,\ldots,i-1$, we can define w'_i by

$$(51) \qquad B._{\beta_i} w'_i(t) + K._{\beta_i} \int_{t'_{i-1}}^{t} w'_i(s)ds = b(t) - K \int_{0}^{t'_{i-1}} x'(s)ds, \ t \geq t'_{i-1}.$$

In attempting to find t'_i we shall try to satisfy the following:

(i) $t'_i \to t_i$ as $\epsilon \to 0$.

(ii) w'_{i+1} may not contain δ's.

(iii) On neighborhoods to the left and right of t'_i, we must have $w'_i(t) \geq 0$ and $w'_{i+1}(t) \geq 0$ respectively.

(iv) w'_{i+1} must be 'close' to w_{i+1} in order that future breakpoints can be adjusted.

We begin by considering $w'_1(t) - w_1(t)$ *for* $t \geq t'_0$. From (17) and the construction of θ, it is clear that

$$(52) \qquad w'_1(t) - w_1(t) = \int_0^{\epsilon} \psi(t-s)\ \theta(s)ds .$$

The following lemma gives an important estimate.

(53) <u>Lemma</u>. There is an $M > 0$, independent of ϵ, such that

$$\int_0^{\epsilon} \theta(s)ds = M\epsilon^{q+1} + O(\epsilon^{q+2})$$

where q is as in the previous section.

Proof. From (39) we have that

$$\theta(t) = \rho_1 \frac{t^q}{q!} + \frac{(\rho_1 + \rho_2)}{q!} \sum_{k=1}^{i-1} (-1)^k (t - \tau_k \epsilon)^q + O(\epsilon^{q+1})$$

for $t \in [\tau_{i-1}\epsilon, \tau_i \epsilon)$. Integrating over $[0,\epsilon)$ and simplifying yields

$$\int_0^{\epsilon} \theta(t)dt = (\rho_1 + \rho_2) \left\{ \frac{\rho_1}{\rho_1 + \rho_2} - \sum_{k=1}^{q} (-1)^{k+1} (1-\tau_k)^{q+1} \right\} \frac{\epsilon^{q+1}}{(q+1)!} + O(\epsilon^{q+2}) .$$

By Lemma D(3) and Theorem (34) the term in { } is equal to

$$\frac{\rho_1}{\rho_1 + \rho_2} \prod_{k=1}^{q} (1 - z_k) + O(\epsilon)$$

where the z_k satisfy (45) and are independent of ϵ. Thus

$$\int_0^{\epsilon} \theta(t)dt = \rho_1 \prod_{k=1}^{q} (1 - z_k) \frac{\epsilon^{q+1}}{(q + 1)!} + O(\epsilon^{q+2}) .$$

Since $1 > z_1 > \cdots > z_q > 0$ and $\rho_1 > 0$, setting

$$M = \rho_1 \prod_{k=1}^{q} (1 - z_k)/(q + 1)!$$

gives the desired result. □

(54) Corollary. Let $g:[0,\infty) \to R$ be such that for all $t \in [0,\epsilon)$

$$g(t) = g(0) + O(\epsilon) .$$

Then there is an $M > 0$ such that

$$\int_0^{\epsilon} g(s)\ \theta(s)ds = g(0)\ M\epsilon^{q+1} + O(\epsilon^{q+2})\ .$$

Proof. This follows immediately from the Lemma. □

We apply this corollary to (52). Since w_1', w_1 and ψ are analytic, differentiating (52) yields

$$(w_1' - w_1)^{(j)}(t) = \int_0^{\epsilon} \psi^{(j)}(t-s)\ \theta(s)ds, \qquad j \geq 0\ .$$

By Corollary (54) we obtain, for $i = 1$,

(55) $$(w_i' - w_i)^{(j)}(t) = O(\epsilon^{q+1}), \qquad j \geq 0\ .$$

Thus not only are w_1 and w_1' close together but so too are all their higher order derivatives. It will be important that the pairs w_i, w_i', $i \geq 1$ inherit this property as well as the same order of magnitude difference, ϵ^{q+1}.

Let us now make the induction hypotheses that $t_{i-1}' \to t_{i-1}$ as $\epsilon \to 0$ and that w_i and w_i' satisfy (55).

Under nondegeneracy, we have that $w_i(t) > 0$ for $t \in (t_{i-1}, t_i)$. Preassign any small $\delta > 0$. Then by this remark and the induction hypotheses there is a $\gamma > 0$ such that for all ϵ sufficiently small,

(56) $$\xi_i'(t) \geq \gamma \qquad \text{for} \quad t \in [t_{i-1} + \delta,\ t_i - \delta]\ ,$$

(57) $$|\xi_i'(t)| \geq \gamma \qquad \text{for} \quad t \in [t_i + \delta,\ t_i + 2\delta]\ ,$$

and all other components of w_i' are $\geq \gamma$ for $t \in [t_{i-1} + \delta, t_i + 2\delta]$. Relation (57) follows from the fact that since ξ_i is analytic, it has at most one zero in $[t_i - \delta, t_i + \delta]$, and is nonzero on $[t_i + \delta, t_i + 2\delta]$.

In Figure 6 below we illustrate the possible cases. The dotted curves below are $\xi_i(\cdot)$ and the solid curves are $\xi_i'(\cdot)$.

It is clear from these figures that a good candidate for t_i' is either $t_i' = t_i$ if ξ_i' has no zero in $[t_i - \delta, t_i + \delta]$, or t_i' equal to the leftmost zero of ξ_i' in $[t_i - \delta, t_i + \delta]$. We shall show that this is indeed the case whenever adjustment of t_i is possible. What determines this is whether or not η_i' and ξ_i' can be related analogously to (48).

The following is the main result.

(58) Proposition. Adjusting t_i.

Assume the induction hypotheses that $t_{i-1}' \to t_{i-1}$ and that w_i' and w_i satisfy (55).

From the above discussion we have that for all ϵ sufficiently small, the number of zeros of ξ_i' in $[t_i - \delta, t_i + \delta]$ is constant. Determine t_i' as follows: if there are no zeros in this interval, let $t_i' = t_i$; else let t_i' be the leftmost such zero. For this t_i' define w_{i+1}' as in (51). In addition let $p_i' \geq 0$ be such that

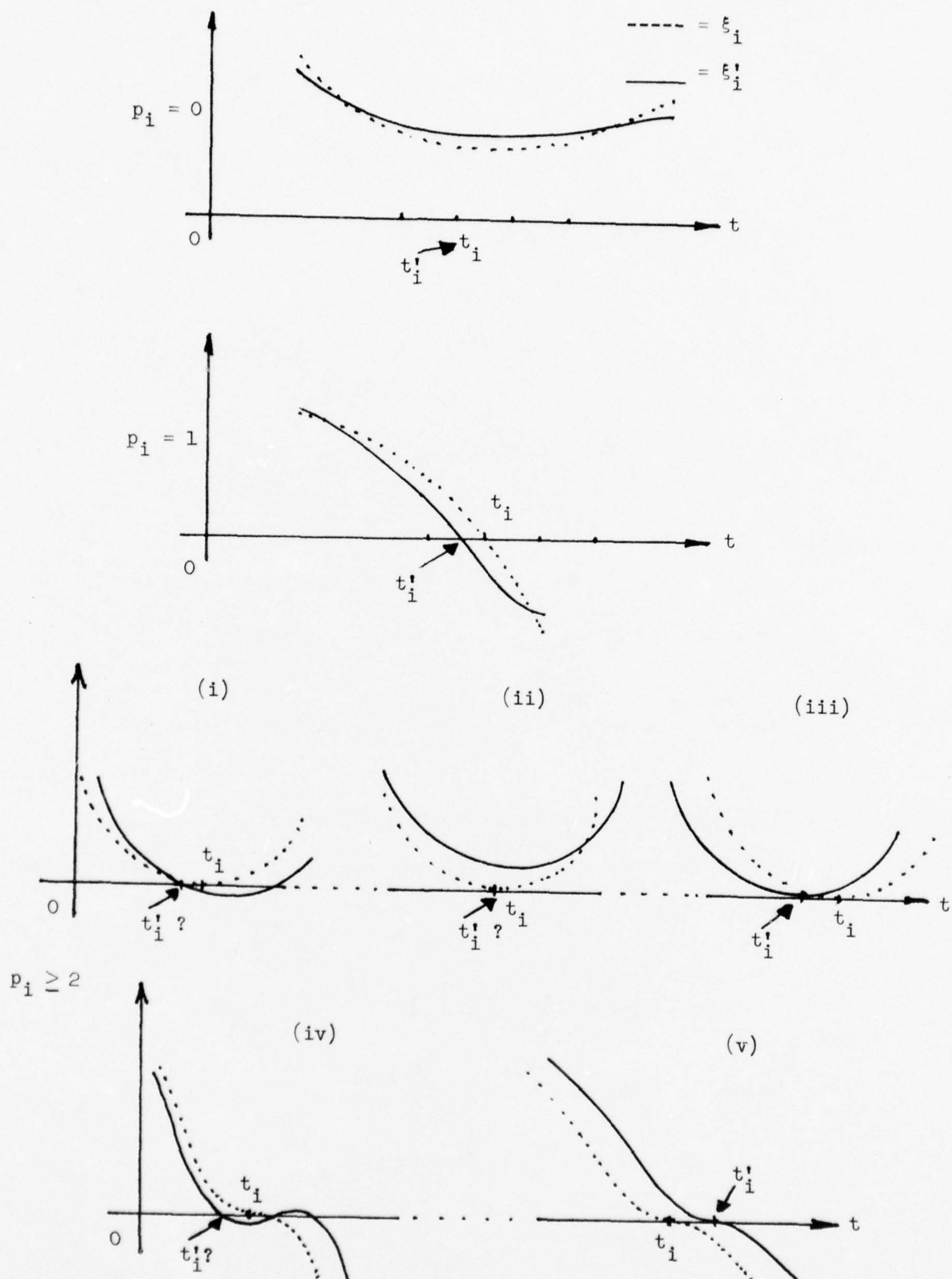


FIGURE 6. ξ_i and ξ_i'

$$\xi_i'^{(j)}(t_i') = 0\ , \qquad 0 \leq j < p_i'$$
$$\xi_i'^{(p_i')}(t_i') \neq 0\ .$$

Then, <u>if</u> $p_i' = p_i$, the following hold:

(i) $t_i' - t_i = O(\epsilon^{q+1})$ if $p_i > 0$

(ii) $(w_{i+1}' - w_{i+1})^{(j)}(t) = O(\epsilon^{q+1})$ for all $t \geq \max\{t_i', t_i\}$ and $j \geq 0$

(iii) $w_{i+1}'(t) \geq 0$ for $t \in [t_i', t_{i+1} - \delta]$.

<u>Proof</u>.

(i) Write $\xi_i'(t) = \xi_i(t) + y(t)$. Differentiating this p_i-1 times and expanding $\xi_i^{(p_i-1)}$ about t_i yields

$$\xi_i'^{(p_i-1)}(t) = y^{(p_i-1)}(t) + \xi_i^{(p_i)}(t_i)(t-t_i) + O((t-t_i)^2)$$

Now from (55), (56), and (57) it follows that $t_i' \to t_i$ as $\epsilon \to 0$. Further $\xi_i'^{(p_i-1)}(t_i') = 0$ and $\xi_i^{(p_i)}(t_i)$ is independent of ϵ. Thus

$$\frac{y^{(p_i-1)}(t_i')}{(t_i' - t_i)} \to -\xi_i^{(p_i)}(t_i) \neq 0 \qquad \text{as } \epsilon \to 0\ .$$

Since, by (55), $y^{(p_i-1)}(t_i') = O(\epsilon^{q+1})$,

$t_i' - t_i = O(\epsilon^{q+1})$ as required.

(ii) Since $p_i' = p_i$, it is clear from Lemma (47) that ξ_i' and η_i' satisfy

$$\sum_{k=0}^{q_i} v_k^i \, \eta_i'^{(k)}(t) + \int_{t_i'}^{t} \varphi_i(t-s)\, \eta_i'(s)ds = \xi_i'^{(p_i)}(t) , \qquad t \geq t_i' \tag{59}$$

and

$$\eta'^{(k)}(t_i') = 0, \qquad 0 \leq k < q_i$$

$$\eta'^{(q_i)}(t_i') \neq 0 . \tag{60}$$

We wish to solve for η_i and η_i' in terms of ξ_i and ξ_i', respectively. Since the equations (48) and (59) have the same coefficients, and since both η_i and η_i' have their first q_i-1 derivatives initially zero, it is easily seen (by using Laplace transforms, for example) that

$$\eta_i(t) = \tilde{v}_i \xi_i^{(p_i)}(t) + \int_{t_i}^{t} \tilde{\varphi}_i(t-s)\, \xi_i^{(p_i)}(s)ds$$

and

$$\eta_i'(t) = \tilde{v}_i \xi_i'^{(p_i)}(t) + \int_{t_i'}^{t} \tilde{\varphi}_i(t-s)\, \xi_i'^{(p_i)}(s)ds$$

for some $\tilde{v}_i$ and $\tilde{\varphi}_i$.[1] Differentiating these yields for $j \geq 0$

$$\begin{aligned}(61) \quad & (\eta_i' - \eta_i)^{(j)}(t) \\ & = \tilde{v}_i(\xi_i' - \xi_i)^{(p_i+j)}(t) + \sum_{k=0}^{j-1} \tilde{\varphi}_i^{(k)}(0)(\xi_i' - \xi_i)^{(p_i+j-1-k)}(t) \\ & \quad + \int_{t_i^*}^{t} \tilde{\varphi}_i^{(j)}(t-s)(\xi_i' - \xi_i)^{(p_i)}(s)ds + \Omega_i(t)\end{aligned}$$

[1] $\tilde{v}_i = 0$ iff $q_i > 0$.

for all $t \geq t_i^*$, where $t_i^* = \max\{t_i', t_i\}$ and

$$\Omega_i(t) = -\int_{t_i}^{t_i'} \tilde{\varphi}_i^{(j)}(t-s)\, \xi_i^{(p_i)}(s)ds \quad \text{if} \quad t_i' \geq t_i$$

and

$$\Omega_i(t) = \int_{t_i'}^{t_i} \tilde{\varphi}_i^{(j)}(t-s)\, \xi_i'^{(p_i)}(s)ds \quad \text{if} \quad t_i' \leq t_i \ .$$

Since $t_i' - t_i = O(\epsilon^{q+1})$ and $\tilde{\varphi}_i^{(j)}(\cdot)$, $\xi_i^{(p_i)}(\cdot)$ and $\xi_i'^{(p_i)}(\cdot)$ are bounded, we have

$$\Omega_i(t) = O(\epsilon^{q+1}) \ .$$

By (55) we also have that all the remaining terms of (61) are $O(\epsilon^{q+1})$. Hence

(62) $$(\eta_i' - \eta_i)^{(j)}(t) = O(\epsilon^{q+1}), \ j \geq 0, \ t \geq t_i^* \ .$$

Now since the components of w_{i+1}, except η_i, are all components of x_{β_i} for $t \in [t_i, t_{i+1})$, we see from (44) that for $t \geq t_i^*$,

$$(w_{i+1}' - w_{i+1})_k(t) = (w_i' - w_i)_k(t) + \text{finitely many terms in } (\eta_i' - \eta_i)^{(j)}(t)$$

where k runs over all components of w_{i+1} except for η_i. By (62) and (55) this implies that

$$(w'_{i+1} - w_{i+1})^{(j)}(t) = O(\epsilon^{q+1})$$

for $t \geq t_i^*$, $j \geq 0$, as required for (ii).

(iii) This follows immediately from (i) and (ii) and since $\eta_i^{(q_i)}(t_i) > 0$. $\square$

Since the induction hypotheses of the Proposition were shown to be true for $i = 1$ we can now inductively perturb the breakpoints t_i provided at each step we have $p_i' = p_i$. When $p_i = 0$ or $p_i = 1$, it can be easily seen that this is always the case. However, when $p_i \geq 2$, we only obtain $p_i' = p_i$ in the cases (iii) and (v) of Figure 5. The remaining cases in Figure 5 for $p_i \geq 2$ all have $p_i' < p_i$. When this occurs, statements (i), (ii) and (iii) of the Proposition need no longer hold. This may result in infeasibilities or in very large perturbations further on. As we shall see in the next section, statements,(i) and (ii) are crucial in establishing improvement in the objective.

The only breakpoint requiring special attention is the endpoint $t_k = T$. If the effect of θ is to move t_k' to the right, then we simply truncate the new solution at T. However, if t_k' moves to the left, then the attempted basis change will only be possible if there is some feasible extension of x beyond T. As will be indicated in the next section, even when we can move t_k' to the left, we may not be able to obtain a decrease in the objective value.

In the event that we get 'stuck' at the endpoint or at some t_i because $p_i' < p_i$, it may be possible to take advantage of an ambiguity in the dual prices. See Section 4.5.

4.2.3. Proof that x' is an improved solution.

Assuming that the adjustments of the breakpoints on $t \geq \epsilon$ were successful[1] we now have a new basic feasible solution, x', specified by the bases σ_i on $[\tau_{i-1}\epsilon, \tau_i\epsilon)$, and β_i on $[t'_{i-1}, t'_i)$. To show that

$$\int_0^T c(t)\ x'(t)dt < \int_0^T c(t)\ x(t)dt$$

we shall show that the decrease obtained over $[0,\epsilon)$ dominates any possible increase due to the shifting of the breakpoints. Of course, as discussed in Section 3.2, if none of the breakpoints have shifted the result is immediate.

The case where $t'_k < T$ is special and will be considered separately at the end of this section.

The change in the objective may be found using (3) and the fact that complementary slackness, (8), holds for the pair (x,λ). The expression so obtained is

$$(63) \qquad \int_0^T c(t)(x'(t) - x(t))dt = \int_0^T \bar{c}(t)\ x'(t)dt - \sum_i \sum_j \bar{d}^{ji}\ x'^{(j)}(t_i^-)\ .$$

[1] i.e. $p_i' = p_i$, all i, and $t'_k = T$, where $t_k = T$.

We need to show that the right-hand side is negative for ϵ sufficiently small. Define

$$h_1(\epsilon) = \int_0^{\epsilon} \bar{c}(t)\; x'(t)dt$$

$$h_2(\epsilon) = \int_{\epsilon}^{T} \bar{c}(t)\; x'(t)dt - \sum_i \sum_j \bar{d}^{ji}\; x'^{(j)}(t_i^-)\;.$$

(64) Lemma. There is an $M < 0$ such that

$$h_1(\epsilon) = M\epsilon^{q+1} + O(\epsilon^{q+2}) \quad .$$

Proof. By complementary slackness

$$h_1(\epsilon) = \int_0^{\epsilon} \bar{c}_{\ell}(t)\; \theta(t)dt \;.$$

Since $\bar{c}_{\ell}(\cdot)$ is analytic on a neighborhood of $t = 0$, and $\bar{c}_{\ell}(0) < 0$, we can apply Corollary (54) and obtain the result. □

We next show that $h_2(\epsilon) = O(\epsilon^{q+2})$. By complementary slackness and Proposition (17), we can break down this expression into contributions due to t_i' moving left or right. We obtain

$$(65) \qquad h_2(\epsilon) = \sum_{\substack{i \text{ such that} \\ t_i' < t_i}} \left(\int_{t_i'}^{t_i} \bar{c}_{\ell_i}(t)\; \eta_i'(t)dt - \sum_j \bar{d}_{\ell_i}^{ji}\; \eta_i'^{(j)}(t_i)\right)$$

$$+ \sum_{\substack{i \text{ such that} \\ t_i < t_i'}} \int_{t_i}^{t_i'} \bar{c}_{r_i}(t)\; \xi_i'(t)dt \;.$$

Note that the terms in $t_i' = t_i$ make no contribution. Thus for all i appearing in (65), we have, from the construction of the t_i', that $p_i > 0$.

(66) Lemma. For all i such that $t_i < t_i'$,

$$\int_{t_i}^{t_i'} \bar{c}_{r_i}(t)\, \xi_i'(t)dt = O(\epsilon^{q+2}) .$$

Proof. $\xi_i'(t) = (\xi_i'(t) - \xi_i(t)) + \xi_i(t)$. By Proposition (58)

$$\xi_i'(t) - \xi_i(t) = O(\epsilon^{q+1}) .$$

Expanding ξ_i about t_i yields

$$\xi_i(t) = O((t - t_i)^{p_i}) .$$

By Proposition (58)

$$t_i' - t_i = O(\epsilon^{q+1}) .$$

Since $p_i > 0$, combining the above establishes the Lemma. □

Before continuing we need to establish an important connection between the reduced costs immediately before and after the basis change at t_i, $t_i < T$.

(67) Lemma. Let p be such that $(\beta_i)_p = r_i$. Let $\bar{n}_i \geq 0$ be such that

$$\bar{c}_{r_i}^{(j)}(t_i^+) = 0\ , \qquad 0 \leq j < \bar{n}_i$$

$$\bar{c}_{r_i}^{(\bar{n}_i)}(t_i^+) \neq 0\ .^{1}$$

Let $\hat{h}_\ell(s)$ be defined by

$$\hat{h}_\ell(s) = (B._{\beta_i} + \frac{1}{s} K._{\beta_i})^{-1} (b_\ell + \frac{1}{s} k_\ell)$$

and let ω_ℓ and n_ℓ be such that the leading term in the Laurent expansion of $\hat{h}_{\ell p}(s)$ is $\omega_\ell s^{n_\ell}$. Then

(i) $\bar{d}_\ell^{ji} = 0$ for $j \geq \max\{0, n_\ell - \bar{n}_i\}$

(ii) if $n_\ell - \bar{n}_i > 0$, $\bar{d}_\ell^{n_\ell - \bar{n}_i - 1, i} = (-1)^{\bar{n}_i} \omega_\ell \bar{c}_{r_i}^{(\bar{n}_i)}(t_i^+)$

(iii) if $n_\ell - \bar{n}_i \leq 0$,

$$\bar{c}_\ell^{(j)}(t_i^-) = \bar{c}_\ell^{(j)}(t_i^+), \qquad 0 \leq j < \bar{n}_i - n_\ell$$

and

$$\bar{c}_\ell^{(\bar{n}_i - n_\ell)}(t_i^-) = \bar{c}_\ell^{(\bar{n}_i - n_\ell)}(t_i^+) - (-1)^{n_\ell} \omega_\ell \bar{c}_{r_i}^{(\bar{n}_i)}(t_i^+)\ .$$

---

[1] If no such $\bar{n}_i$ exists then $\bar{c}_{r_i}(t)$ is identically zero to the right of t_i. Set $\bar{n}_i = 0$.

Proof. We shall use Laplace Transforms. By (11), λ satisfies[1]

$$\lambda(t)\, B_{\cdot\beta_i} + \int_t^{t_i} \lambda(s)ds\, K_{\cdot\beta_i} = c_{\beta_i}(t) - \int_{t_i}^{T} \lambda(s)ds\, K_{\cdot\beta_i}$$

for $t \in (t_{i-1}, t_i)$ in the distributional sense. For convenience we shall reverse the time direction and shift the time origin to t_i. Define μ on $[0,\infty)$ by

$$\mu(t)\, B_{\cdot\beta_i} + \int_0^{t} \mu(s)ds\, K_{\cdot\beta_i} = g_{\beta_i}(t), \qquad t \geq 0 \tag{68}$$

where

$$g(t) = c(t_i - t) - \int_{t_i}^{T} \lambda(s)ds\, K$$

and we assume that $c(\cdot)$ has an analytic extension from $[0,\infty)$ to $(-\infty,\infty)$. Clearly

$$\mu(t_i - t) = \lambda(t) \quad \text{for} \quad t \in (t_{i-1}, t_i)\ .$$

Further, it is easily seen by definition that $-\bar{d}_\ell^{ji}$ is the coefficient of $\delta^{(j)}$ in $\bar{g}_\ell(\cdot)$, defined by

$$\bar{g}_\ell(t) = g_\ell(t) - \mu(t)b_\ell - \int_0^{t} \mu(s)ds\, k_\ell\ , \tag{69}$$

and that $\bar{c}_\ell(t)$ is the functional part of $\bar{g}_\ell(t_i - t)$.

---

[1] Strictly speaking we should write

$$\int_{t_i}^{T} \lambda(s)ds \quad \text{as} \quad \int_{t_i}^{T} \lambda^*(s)ds + \sum_{k>i} \nu^{0k}.$$

Taking Laplace transforms of (68) yields

$$\hat{\mu}(s) = \hat{g}_{\beta_i}(s)\ (B._{\beta_i} + \frac{1}{s} K._{\beta_i})^{-1}$$

and substituting into (69) yields

$$\hat{\bar{g}}_\ell(s) = \hat{g}_\ell(s) - \hat{g}_{\beta_i}(s)\ \hat{h}_\ell(s)\ . \tag{70}$$

We now wish to relate $\hat{\bar{g}}_\ell(s)$ to the reduced costs taken with respect to the basis β_{i+1}. Denote these with the superscript '*'. Thus we define

$$\hat{\bar{g}}^*_\ell(s) = \hat{g}_\ell(s) - \hat{g}_{\beta_{i+1}}(s)\ \hat{h}^*_\ell(s)$$

where

$$\hat{h}^*_\ell(s) = (B._{\beta_{i+1}} + \frac{1}{s} K._{\beta_{i+1}})^{-1}\ (b_\ell + \frac{1}{s} k_\ell)\ .$$

We shall show that

$$\hat{\bar{g}}_\ell(s) = \hat{\bar{g}}^*_\ell(s) - \hat{\bar{g}}^*_{r_i}(s)\ \hat{h}_{\ell p}(s)\ . \tag{71}$$

Note first that $\hat{h}_\ell(s)$ and $\hat{h}^*_\ell(s)$ are related by

$$\begin{aligned} &\hat{h}_{\ell j}(s) = \hat{h}^*_{\ell j}(s) - \hat{h}^*_{r_i j}(s)\ \hat{h}^*_{\ell p}(s)/\hat{h}^*_{r_i p}(s), \qquad j \neq p \\ &\hat{h}_{\ell p}(s) = \hat{h}^*_{\ell p}(s)/\hat{h}^*_{r_i p}(s)\ . \end{aligned} \tag{72}$$

This can be easily shown by relating $(B._{\beta_i} + \frac{1}{s} K._{\beta_i})^{-1}$ and

$(B_{\cdot\beta_{i+1}} + \frac{1}{s} K_{\cdot\beta_{i+1}})^{-1}$. Rewrite (70) as

$$\hat{\bar{g}}_\ell(s) = \hat{\bar{g}}^*_\ell(s) + \hat{g}_{\beta_{i+1}}(s)\ \hat{h}^*_\ell(s) - \hat{g}_{\beta_i}(s)\ \hat{h}_\ell(s)$$

and $\hat{g}_{\beta_i}(s)\ \hat{h}_\ell(s)$ as

$$\hat{g}_{\beta_i}(s)\ \hat{h}_\ell(s) = \hat{g}_{\beta_{i+1}}(s)\ \hat{h}_\ell(s) + (\hat{g}_{r_i}(s) - \hat{g}_{\ell_i}(s))\ \hat{h}_{\ell p}(s)\ .$$

Combining these two yields

$$\text{(73)} \qquad \hat{\bar{g}}_\ell(s) = \hat{\bar{g}}^*_\ell(s) + (\hat{g}_{\ell_i}(s) - \hat{g}_{r_i}(s))\ \hat{h}_{\ell p}(s) + \hat{g}_{\beta_{i+1}}(s)(\hat{h}^*_\ell(s) - \hat{h}_\ell(s)).$$

From (72) we have that

$$\hat{h}^*_{\ell j}(s) - \hat{h}_{\ell j}(s) = \hat{h}^*_{r_i j}(s)\ \hat{h}_{\ell p}(s), \qquad j \neq p$$

$$\hat{h}^*_{\ell p}(s) - \hat{h}_{\ell p}(s) = (\hat{h}^*_{r_i p}(s) - 1)\ \hat{h}_{\ell p}(s)\ .$$

Substituting this into (73) and rearranging yields (71).

From (71) we can now establish the lemma. Observe that $\bar{g}^*_\ell(t_i - t)$, $t \leq t_i$, is the reduced cost $\bar{c}_\ell(t)$, $t \leq t_i$, <u>had there been no basis change at</u> t_i. Hence $\bar{g}^*_\ell(t_i - t)$ is an analytic function and we obtain

$$(-1)^j\ \bar{g}^{*(j)}_\ell(0^+) = \bar{c}^{(j)}_\ell(t_i^+)\ .$$

Applying Lemma C.4(19) to (71) yields

$$\hat{\bar{g}}_\ell(s) = \sum_{j=0}^{\infty} (-1)^j \bar{c}_\ell^{(j)}(t_i^+)/s^{j+1} - (\sum_{j=0}^{\infty} (-1)^j \bar{c}_{r_i}^{(j)}(t_i^+)/s^{j+1})(\omega_\ell s^{n_\ell}+\cdots)$$

Inspecting this expression we see that there are no terms in s^j for $j \geq \max\{0, n_\ell - \bar{n}_i\}$. Hence $\bar{d}_\ell^{ji} = 0$ for such j, and (i) is established. If $n_\ell - \bar{n}_i > 0$, the coefficient of $s^{n_\ell - n_i - 1}$ is $(-1)^{\bar{n}_i+1} \omega_\ell \bar{c}_{r_i}^{(\bar{n}_i)}(t_i^+)$, establishing (ii). If $n_\ell - \bar{n}_i \leq 0$, the coefficient of s^{-j-1} is $(-1)^{(j)} \bar{c}_\ell^{(j)}(t_i^+)$ for $0 \leq j < \bar{n}_i - n_\ell$, so that $\bar{c}_\ell^{(j)}(t_i^-) = \bar{c}_\ell^{(j)}(t_i^+)$.

Further, the coefficient of $s^{-(\bar{n}_i - n_\ell)-1}$ is

$$(-1)^{\bar{n}_i - n_\ell} \bar{c}_\ell^{(\bar{n}_i - n_\ell)}(t_i^+) - (-1)^{\bar{n}_i} \omega_\ell \bar{c}_{r_i}^{(\bar{n}_i)}(t_i^+)$$

so that

$$\bar{c}_\ell^{(\bar{n}_i - n_\ell)}(t_i) = \bar{c}_\ell^{(\bar{n}_i - n_\ell)}(t_i^+) - (-1)^{n_\ell} \omega_\ell \bar{c}_{r_i}^{(n_i)}(t_i^+)$$

as required for (iii). □

(74) <u>Corollary</u>.

(i) $\bar{d}_{\ell_i}^{ji} = 0$ for $j \geq \max\{0, q_i - p_i - \bar{n}_i\}$

(ii) if $q_i - p_i - \bar{n}_i > 0$,

$$\bar{d}_{\ell_i}^{q_i - p_i - \bar{n}_i - 1, i} = (-1)^{\bar{n}_i} v_{q_i}^i \bar{c}_{r_i}^{(\bar{n}_i)}(t_i^+)$$

(iii) if $q_i - p_i - \bar{n}_i \leq 0$,

$$\bar{c}_{\ell_i}^{(j)}(t_i^-) = 0 , \qquad 0 \leq j < \bar{n}_i + p_i - q_i$$

and

$$\bar{c}_{\ell_i}^{(\bar{n}_i+p_i-q_i)}(t_i^-) = (-1)^{q_i-p_i+1} v_{q_i}^i \bar{c}_{r_i}^{(\bar{n}_i)}(t_i^+)$$

where $v_{q_i}^i$ is given in Lemma (47).

<u>Proof</u>. From the proof of Lemma (47) it can be easily shown that $n_{\ell_i} = q_i - p_i$ and that $\omega_{\ell_i} = v_{q_i}^i$. Since x_{ℓ_i} is basic on $[t_i, t_{i+1})$, it follows by complementary slackness that $\bar{c}_{\ell_i}(t) = 0$, $t \in [t_i, t_{i+1})$. Hence $\bar{c}_{\ell_i}^{(j)}(t_i^+) = 0$ for $j \geq 0$. Applying the lemma yields the result. □

We now continue with our proof that $h_2(\epsilon) = O(\epsilon^{q+2})$.

(75) <u>Lemma</u>. For all i such that $t_i' < t_i$

$$\int_{t_i'}^{t_i} \bar{c}_{\ell_i}(t)\, \eta_i'(t) dt - \sum_j \bar{d}_{\ell_i}^{ji}\, \eta_i'^{(j)}(t_i) = O(\epsilon^{q+2}).$$

<u>Proof</u>. Since $t_i' \to t_i$, and since

$$\eta_i'^{(q_i)}(t_i') \to \eta_i^{(q_i)}(t_i) \neq 0 \qquad \text{as} \quad \epsilon \to 0 ,$$

we may write, by (60),

$$\eta_i'^{(j)}(t_i) = O((t_i' - t_i)^{q_i - j}), \qquad 0 \leq j \leq q_i .$$

Since $t_i' - t_i = O(\epsilon^{q+1})$ we obtain

$$\eta_i'^{(j)}(t_i) = O(\epsilon^{(q+1)(q_i-j)}) .$$

Thus

$$\eta_i'^{(j)}(t_i) = O(\epsilon^{q+2}) \qquad \text{when } 0 \leq j \leq q_i - 2 .$$

By Corollary (74)(i), since $p_i > 0$,

$$\bar{d}^{ji}_{\ell_i} = 0 \quad \text{for} \quad j \geq \max\{q_i - 1, 0\}.$$

Thus

$$\sum_j \bar{d}^{ji}_{\ell_i} \eta_i'^{(j)}(t_i) = O(\epsilon^{q+2}) \quad .$$

We next show that

$$\int_{t_i'}^{t_i} \bar{c}_{\ell_i}(t)\ \eta_i'(t)dt = O(\epsilon^{q+2}) \ .$$

If $q_i \geq 1$ it follows since $t_i' - t_i = O(\epsilon^{q+1})$ and since $\eta_i'(t) = O((t_i' - t_i)^{q_i})$. Suppose that $q_i = 0$. Then since $p_i > 0$, we have by Corollary (74)(iii) that $\bar{c}_{\ell_i}(t) = O(t_i' - t_i)$. Since $t_i' - t_i = O(\epsilon^{q+1})$, we are done.

This completes the proof. □

Combining Lemmas (66) and (75) with (65) shows that $h_2(\epsilon) = O(\epsilon^{q+2})$.

We now state our main result.

(76) Theorem. Decrease in the objective

Let x' be constructed as in Sections 4.2.1 and 4.2.2. Then there is an $M < 0$ such that

$$\int_0^T c(t)\, x'(t)dt - \int_0^T c(t)\, x(t)dt = M\epsilon^{q+1} + O(\epsilon^{q+2}) .$$

Proof. This is immediate from the above and Lemma (64). □

In concluding this section we consider the case $t_k' < T$. This can only occur if some $x_{r_k}(T) = 0$, $r_k \in \beta_k$, and x has a feasible extension on a neighborhood to the right of T. Let ℓ_k and q_k refer to the new basic variable beyond T, in keeping with ℓ_i and q_i at the other breakpoints.

The reason why this is special is that Lemma (67) does not apply at the endpoint $t_k = T$. In the event that $q_k \geq 1$, this means that we may have $\bar{d}_{\ell_k}^{jk} \neq 0$ for $j = q_k - 1$, and therefore that Lemma (75) is no longer generally true. Following the proof of Lemma (75), we see that this could yield a positive contribution of $O(\epsilon^{q+1})$ to the objective value if $\bar{d}_{\ell_k}^{q_k-1,k} < 0$. If $q_k = 0$, we may have $\bar{c}_{\ell_k}(T) > 0$, with the same effect.

From an algorithmic point of view, when $\bar{d}_{\ell_k}^{q_k-1,k} < 0$ if $q_k \geq 1$ or $\bar{c}_{\ell_k}(T) > 0$ if $q_k = 0$, we suggest first finding an alternative extension beyond T for which these do not hold. Should this not be possible, simply allow t_k' to adjust. If a net increase has been obtained then the whole iteration must be disregarded and we either select some other interval of perturbation or use the technique of Section 4.5.

Note that the question of adjusting t_k will in any case only occur at isolated iterations.

4.3. Improving the Solution by Adjusting a Breakpoint

This section considers the adjustment of some breakpoint t_i to obtain an improved solution. As in Section 4.2 we begin with a given basic solution x and its associated reduced costs $\bar{c}(\cdot)$ and $\{\bar{d}^{ji}\}$. The only case we consider here is that of $p_i = 0$, i.e. $x_{r_i}(t_i^-) > 0$. The reason for this is that adjusting t_i when $x_{r_i}(t_i^-) = 0$ (i.e. $p_i > 0$) usually causes the occurrence infeasibilities, or changes on $t > t_i$ that have greater orders of magnitude than the adjustment at t_i.[1]

That an improvement can be obtained by adjusting t_i will be signaled either by one of $\bar{c}_{r_i}(t_i^+) < 0$, $\bar{c}_{\ell_i}(t_i^-) < 0$, or by some $\bar{d}^{ji}_{\ell_i} \neq 0$. While the former is a special case of that considered in in Section 4.2, it is much more convenient to think only in terms of adjusting a breakpoint.

Throughout this section i will be held fixed, and $0 < t_i < T$. We shall let t_i' be one of $t_i - \epsilon$, $t_i + \epsilon$ for $\epsilon > 0$ and small, and then adjust the remaining breakpoints $t_r > t_i$ in the same way as was done in Section 4.2.2. All notation will be as in the previous section.

[1] Cf. Remark after Proposition (58).

AD-A065 771 STANFORD UNIV CALIF SYSTEMS OPTIMIZATION LAB F/G 12/1
FUNDAMENTALS OF A CONTINUOUS TIME SIMPLEX METHOD.(U)
DEC 78 A F PEROLD N00014-75-C-0267
UNCLASSIFIED SOL-78-26 NL

2 OF 2
AD
A065771

END
DATE
FILMED
5-79
DDC

Observe that setting $t_i' = t_i - \epsilon$ makes x_{ℓ_i} basic over $[t_i', t_i)$ and that setting $t_i' = t_i + \epsilon$ makes x_{r_i} basic over $[t_i, t_i')$. As in (65) we can obtain the local contributions to the objective, attributable to this change, in terms of $\bar{c}$, $\{\bar{d}^{ij}\}$, $\xi_i(\cdot)$ and $\eta_i(\cdot)$. Let $g_L(\epsilon)$ and $g_R(\epsilon)$ respectively denote the contributions due to moving left and right. These may be written as

$$g_L(\epsilon) = \int_{t_i'}^{t_i} \bar{c}_{\ell_i}(t)\, \eta_i'(t)dt - \sum_j \bar{d}_{\ell_i}^{ji}\, \eta_i'^{(j)}(t_i) \tag{77}$$

and

$$g_R(\epsilon) = \int_{t_i}^{t_i'} \bar{c}_{r_i}(t)\, \xi_i(t)dt \tag{78}$$

We illustrate these two cases (Figure 7 below) in the event that $q_i > 0$. When $q_i = 0$, η and η' will look like ξ_i in Figure 6 with the time direction reversed.

The following lemma relates $g_L(\epsilon)$ and $g_R(\epsilon)$.

(79) <u>Lemma</u>. Let $\bar{n}_i$ be as in Lemma (67).

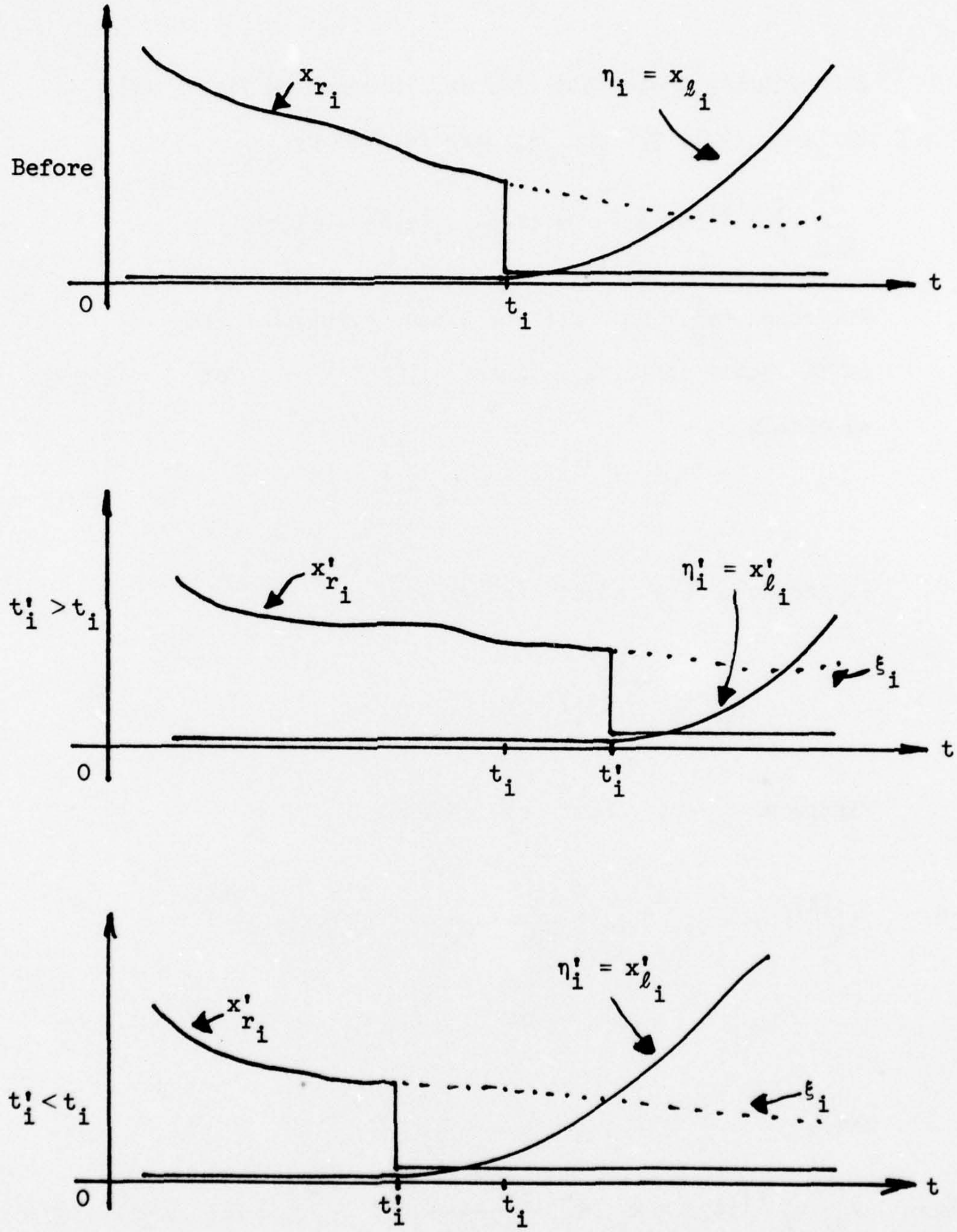
(i) $g_R(\epsilon) = \bar{c}_{r_i}^{(\bar{n}_i)}(t_i^+)\, x_{r_i}(t_i^-) \dfrac{\epsilon^{\bar{n}_i+1}}{(\bar{n}_i+1)!} + O(\epsilon^{\bar{n}_i+2})$

(ii) $g_L(\epsilon) = (-1)^{\bar{n}_i+1} g_R(\epsilon) + O(\epsilon^{\bar{n}_i+2})$.

<u>Proof</u>.

(i) By definition of $\bar{n}_i$,

$$\bar{c}_{r_i}(t) = \bar{c}_{r_i}^{(\bar{n}_i)}(t_i^+) \frac{(t-t_i)^{\bar{n}_i}}{\bar{n}_i!} + O((t-t_i)^{\bar{n}_i+1}) .$$



Before
x_{r_i}
$\eta_i = x_{\ell_i}$
0
t_i
t
$t_i' > t_i$
x_{r_i}'
$\eta_i' = x_{\ell_i}'$
ξ_i
0
t_i
t_i'
t
$t_i' < t_i$
x_{r_i}'
$\eta_i' = x_{\ell_i}'$
ξ_i
0
t_i'
t_i
t

FIGURE 7

Also

$$\xi_i(t) = x_{r_i}(t_i^-) + O(t-t_i) \ .$$

Substituting these into (78) and integrating yields (i).

(ii) By Lemma (47), ξ_i and η_i' are related by

$$\sum_{k=0}^{q_i} v_k^i \eta'^{(k)}(t) + \int_{t_i'}^{t} \varphi_i(t-s)\, \eta_i'(s)ds = \xi_i(t)$$

for some $\{v_k^i\}$ and $\varphi_i(\cdot)$. Since $\xi_i(t_i') > 0$ for ϵ sufficiently small, and since $\eta_i'^{(k)}(t_i') = 0$ for $0 \leq k < q_i$, we obtain

$$v_{q_i}^i \eta_i'^{(q_i)}(t_i') = \xi_i(t_i') \ .$$

Expanding $\xi_i(\cdot)$ about t_i yields

(80) $$v_{q_i}^i \, \eta_i'^{(q_i)}(t_i') = x_{r_i}(t_i^-) + O(\epsilon) \ .$$

Expanding $\eta_i'(\cdot)$ about t_i' yields

(81) $$\eta_i'^{(k)}(t) = \eta_i'^{(q_i)}(t_i') \frac{(t-t_i')^{q_i-k}}{(q_i-k)!} + O((t-t_i')^{q_i-k+1})$$

$$\text{for } 0 \leq k \leq q_i \ .$$

Hence

(82) $$v_{q_i}^i \, \eta_i'^{(k)}(t_i) = x_{r_i}(t_i^-) \frac{\epsilon^{q_i-k}}{(q_i-k)!} + O(\epsilon^{q_i-k+1}) \quad \text{for } 0 \leq k \leq q_i \ .$$

Next, apply Corollary (74). If $q_i - \bar{n}_i > 0$, then the leading $\bar{d}^{ji}_{\ell_i}$ in (77) is

$$\bar{d}^{q_i-\bar{n}_i-1,i}_{\ell_i} = (-1)^{\bar{n}_i} v^i_{q_i} \bar{c}^{(\bar{n}_i)}_{r_i}(t_i^+) \ .$$

From (82) we then have

$$(83) \quad -\bar{d}^{q_i-\bar{n}_i-1,i}_{\ell_i} \eta_i'^{(q_i-\bar{n}_i-1)}(t_i^+)$$

$$= (-1)^{\bar{n}_i+1} \bar{c}^{(\bar{n}_i)}_{r_i}(t_i^+) \, x_{r_i}(t_i^-) \frac{\epsilon^{\bar{n}_i+1}}{(\bar{n}_i+1)!} + 0(\epsilon^{\bar{n}_i+2}) \ .$$

Also by (82) it is clear that all other terms in (77) are $0(\epsilon^{\bar{n}_i+2})$. Hence $g_L(\epsilon)$ is given by the right hand side of (83).

If $q_i - \bar{n}_i \leq 0$ then all the $\bar{d}^{ji}_{\ell_i} = 0$, and we may write, by Corollary (74),

$$\bar{c}_{\ell_i}(t) = (-1)^{q_i+1} v^i_{q_i} \bar{c}^{(\bar{n}_i)}_{r_i}(t_i^+) \frac{(t-t_i)^{\bar{n}_i-q_i}}{(\bar{n}_i-q)!} + 0\left((t-t_i)^{\bar{n}_i-q_i+1}\right)$$

for $t_i' \leq t \leq t_i$. Multiply this expression for $\bar{c}_{\ell_i}(t)$ by that given for $\eta_i'(t)$ in (81) with $k = 0$. This yields

$$\bar{c}_{\ell_i}(t) \, \eta_i'(t)$$

$$= (-1)^{\bar{n}_i+1} v^i_{q_i} \eta_i'^{(q_i)}(t_i') \, \bar{c}^{(\bar{n}_i)}_{r_i}(t_i^+) \frac{(t-t_i')^{q_i} (t_i-t)^{\bar{n}_i-q_i}}{q_i! \, (\bar{n}_i - q_i)!} + 0(\epsilon^{\bar{n}_i+1})$$

for $t_i' \leq t \leq t_i$.

To evaluate $\int_{t_i'}^{t_i} \bar{c}_{\ell_i}(t)\ \eta_i'(t)dt$ observe that since $t_i' = t_i - \epsilon$, the value of

$$\int_{t_i'}^{t_i} (t-t_i')^{q_i} (t_i - t)^{\bar{n}_i - q_i} dt \quad .$$

is proportional to a beta-function, and indeed evaluates to

$$\frac{q_i!\ (\bar{n}_i - q_i)!}{(\bar{n}_i + 1)!} \epsilon^{\bar{n}_i + 1} \quad .$$

This then yields

$$\int_{t_i'}^{t_i} \bar{c}_{\ell_i}(t)\ \eta_i'(t)dt$$

$$= (-1)^{\bar{n}_i + 1} v^i_{q_i}\ \eta_i'^{(q_i)}(t_i')\ \bar{c}_{r_i}^{(\bar{n}_i)}(t_i^+) \frac{\epsilon^{\bar{n}_i + 1}}{(\bar{n}_i + 1)!} + O(\epsilon^{\bar{n}_i + 2}) \quad .$$

Using (80), we obtain that $g_L(\epsilon)$ is equal to the right-hand side of (83).

We have thus shown that $g_L(\epsilon)$ is given by the right-hand side of (83) for all q_i, $\bar{n}_i$. By (i), (ii) now follows immediately, and the proof is complete. □

This lemma tells us precisely what to expect locally when adjusting the breakpoint, and only requires information about the behavior of $\bar{c}_{r_i}$ at $t = t_i^+$. We summarize our conclusions in Table 2 below.

Table 2. <u>The Local Contribution to the Objective by Moving</u> t_i <u>to</u> $t_i - \epsilon$ <u>and</u> $t_i + \epsilon$.[1]

	$\bar{c}_{r_i}^{(\bar{n}_i)}(t_i^+) > 0$		$\bar{c}_{r_i}^{(\bar{n}_i)}(t_i^+) < 0$	
	$\bar{n}_i$ even	$\bar{n}_i$ odd	$\bar{n}_i$ even	$\bar{n}_i$ odd
$t_i' = t_i - \epsilon$	-	+	+	-
$t_i' = t_i + \epsilon$	+	+	-	-

The '+' and '-' denote increases and decreases of $O(\epsilon^{\bar{n}_i+1})$ respectively, and were obtained from Lemma (79) by noting that $x_{r_i}(t_i^-) > 0$.

The case that we shall be most interested in is that of $\bar{n}_i = 0$, i.e. $\bar{c}_{r_i}(t_i^+) \neq 0$. This is because, as we shall next indicate, it is the only case in which a decrease obtained by the above adjustment always dominates any possible increases due to the adjustment of the time points $t_r > t_i$. However, if, for example, no other breakpoints need be adjusted, then shifting t_i according to Table 2 will yield a decrease for $\bar{n}_i > 0$.

It is interesting to note the analogue of Lemma (79) (with $\bar{n}_i = 0$) in the simplex method: if a variable x_ℓ enters a basis β

[1]Recall that $\bar{n}_i$ is the least integer such that $\bar{c}_{r_i}^{(\bar{n}_i)}(t_i^+) \neq 0$.

replacing x_r and forming a new basis β', then the reduced cost of x_r with respect to β' is opposite in sign to the reduced cost of x_ℓ with respect to β.

To complete the procedure, it remains for us to follow the same course as in Section 4.2.2, and determine whether or not the breakpoints $t_r > t_i$ can be adjusted so as to accommodate the change at t_i and yield an overall decrease. It should also be checked that the adjustment at t_i has caused no immediate infeasibilities, i.e. some other variables becoming negative near t_i.

This can all be done by directly applying the tools and results of Section 4.2, and the details will be omitted. Proposition (58) applies as it stands with $q = 0$; that is, the shifts in the breakpoints, $t_r' - t_r$, are $O(\epsilon)$ and $(w_r' - w_r)^{(j)}(t) = O(\epsilon)$, $r > i$. When $\bar{n}_i = 0$ and we have determined t_i' by Table 2 so as to yield a local decrease, Theorem (76) too applies, with $q = 0$, to yield an overall decrease of $O(\epsilon)$.

4.4. Improving the Solution When Some $\bar{d}_\ell^{ji} \neq 0$

From section 4.1 we see that the $\{\bar{d}_\ell^{ji}\}$ are dual infeasible if for any i, ℓ, and $j \geq 1$

$$\bar{d}_\ell^{0i} > 0 ,$$

$$\text{or} \quad \bar{d}_\ell^{ji} \neq 0 . \quad [1]$$

[1] From now on it will be understood that "$\bar{d}_\ell^{ji} \neq 0$" means $\bar{d}_\ell^{0i} > 0$ when $j = 0$.

In such a case, even though we may have $\bar{c}(t) \geq 0$, $t \in [0,T]$, it may still be possible to improve the solution by making basic, in some suitable way, the variable $x_\ell(t_i^-)$. One such case was considered in Section 4.3 where with $\ell = \ell_i$, it sometimes paid off to move the breakpoint t_i. In general, however, a construction along the lines of that given in Section 4.2.1 is required. Thus, if $\bar{d}_\ell^{ji} \neq 0$, we wish to increase $x_\ell = \theta$ on a neighborhood of t_i in such a way that

$$
(84) \qquad \begin{aligned}
&\theta^{(j)}(t_i) > 0 && \text{if} \quad \bar{d}_\ell^{ji} > 0 , \\
&\theta^{(j)}(t_i^-) < 0 && \text{if} \quad \bar{d}_\ell^{ji} < 0 .
\end{aligned}
$$

This is because $-\bar{d}_\ell^{ji}$ appears as the coefficient of $x_\ell^{(j)}(t_i)$ in (3).

It is important to note that if $\bar{d}_\ell^{ji} \neq 0$, we are able to make $O(1)$ changes in $x_\ell^{(q)}$ only for some $q > j$. This follows from Lemma (67). From this we also see that if we are to make $x_\ell = \theta$ basic over $[t_i - \epsilon_1, t_i + \epsilon_2)$, $\epsilon_1 + \epsilon_2 = \epsilon$, the dominating local contribution to the objective will be given by

$$-\bar{d}_\ell^{j*,i} \, \theta^{(j*)}(t_i)$$

where $j*$ is the largest j such that

$$\bar{d}_\ell^{ji} \neq 0 .$$

Thus we shall pick our perturbation θ such that (84) holds with $j = j*$. See Figure 8 below for an illustration of the case $j* = 1$.

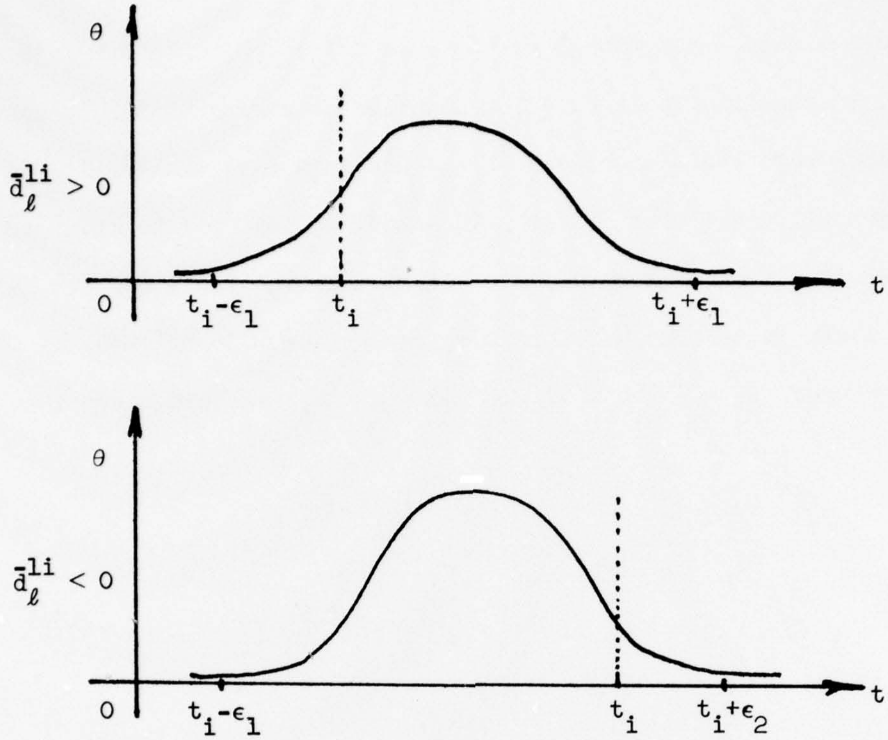


FIGURE 8

In the event that $\ell = \ell_i$ and, say, $\bar{d}^{1i}_{\ell_i} < 0$, with $j^* = 1$, we could obtain a perturbation of the following form:[1]

[1]Since x_{ℓ_i} is basic on $[t_i, t_{i+1})$.

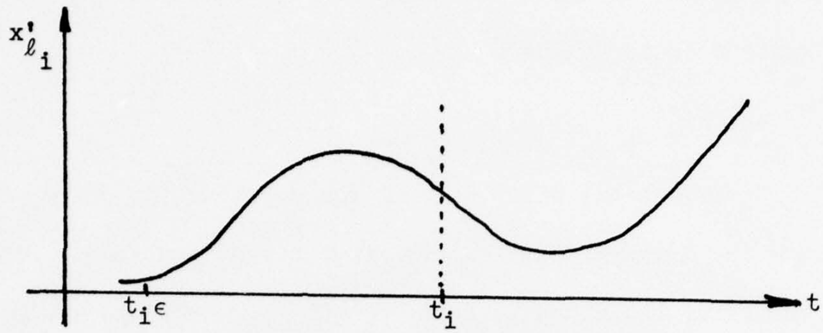


FIGURE 9

We remark that when $\bar{d}_{\ell_i}^{j^*,i} > 0$, a local improvement can always be obtained by just shifting t_i. This follows from Table 2, since by Corollary (74) and $v_{q_i}^i > 0$,

$$\bar{d}_{\ell_i}^{j^*,i} > 0 \Rightarrow (-1)^{\bar{n}_i} \bar{c}_{r_i}^{(\bar{n}_i)}(t_i^+) > 0 .$$

The unfortunate part about in general attempting to improve the solution when some $\bar{d}_{\ell}^{ji} \neq 0$, is that even though we may be able to obtain a local decrease in the objective value, adjusting the breakpoints to the right of this change may sometimes yield an overall increase in the objective value. This was seen to be the case in Section 4.3 with $\bar{n}_i > 0$, or when we needed to adjust $t_k = T$. Further, when $p_i > 0$, we may not even be able to obtain a local improvement in a neighborhood of t_i. However, these are precisely the situations in which the complementary slackness conditions yield alternative dual variables $\lambda^*(\cdot)$ and $\{\nu^{ji}\}$ and which will be considered in the next section.

One particular case where an improvement in the objective value is always possible, is at the endpoint $t_k = T$ when $x_{\beta_k}(T) > 0$. The construction is as follows:

(i) $\bar{d}_\ell^{j*,k} > 0$.

Carry out the first step of the construction of θ in Section 4.2.1 on the interval $[T - \epsilon, T]$, for $\epsilon > 0$ and small. Thus simply make the first basis change from β_k to σ_1 over the whole of this interval. This will yield

$$\theta^{(j*)}(T) = M\epsilon^{q-j*} + O(\epsilon^{q-j*+1})$$

some $M > 0$, and also

$$\theta^{(j)}(T) = O(\epsilon^{q-j}) , \qquad 0 \leq j < j*,$$

where q is determined as in (17). The contribution to the objective value then becomes, using (63),

$$\int_{T-\epsilon}^{T} \bar{c}_\ell(t)\ \theta(t)dt - \sum_{j=0}^{j*} \bar{d}_\ell^{jk}\ \theta^{(j)}(T)$$

$$= -\bar{d}_\ell^{j*,k}\ M\epsilon^{q-j*}/(q - j*)! + O(\epsilon^{q-j*+1})$$

While Lemma (67) does not apply to the endpoint $t_k = T$, it can still be easily shown, by using (70) in the proof of Lemma (67), that $q > j*$. From this it therefore follows that a strict decrease is obtained for ϵ sufficiently small.

(ii) $\bar{d}_{\ell}^{j^*,k} < 0$.

Partition $[T - \epsilon, T]$ into several subintervals over which we alternate the sign of $\theta^{(q)}$, as in Section 4.2.1, in such a way that $\theta^{(j^*)}(T) < 0$. Indeed, by the results of Section 4.2.1, we can choose the intervals so that

$$\theta^{(j^*)}(T) = -M\epsilon^{q-j^*} + O(\epsilon^{q-j^*+1})$$

for some $M > 0$, and also

$$\theta^{(j)}(T) = O(\epsilon^{q-j}), \qquad 0 \le j < j^*.$$

As before, this yields a strict decrease for ϵ sufficiently small.

When $j^* = 1$, the θ so constructed will be as in Figure 8 except with the solution truncated beyond t_i.

At points t_i, $0 < t_i < T$, we remark only that a similar construction applies with the exception that we need to continue beyond t_i to meet requirements of the from (20). In the event that $\ell = \ell_i$, these requirements need not be met, as illustrated in Figure 8.

4.5. Ambiguity in the Dual Variables

In the previous three sections we considered improving our given basic feasible solution when the dual variables, defined by complementary slackness as in (9), were infeasible. However, as was noted there, obtaining an improved solution was not always possible, for the following reasons:

(i) Making a change of basis over an interval of length ϵ may force us to adjust other breakpoints to the right of this interval, and these may violate the condition of Proposition (58), namely $p_i' = p_i$.

(ii) Should this change of basis force $t_k = T$ to move left, it has to be disregarded if there is either no feasible extension of x on a neighborhood to the right of T, or if the result of allowing t_k' to adjust yields an overall increase in the objective value.

(iii) As was noted in Section 4.3 and 4.4, making a change of basis on a neighborhood of some t_i when some $\bar{d}_\ell^{ji} \neq 0$, was in general not possible when $p_i > 0$. Even when we had $p_i = 0$, there were seen to be cases where the need to adjust later time points led to an overall increase in the objective value.

What these cases all have in common is that they are being 'blocked' by some t_i having $p_i > 0$, i.e. the variable $x_{r_i}(t_i^-)$ leaving the basis β_i is zero. This means that the complementary slackness condition

$$x_{r_i}(t_i^-) > 0 \Rightarrow \bar{d}_{r_i}^{0i} = 0 \text{ ,}$$[1]

does not apply, and therefore that the dual variables are no longer uniquely defined.

For such a t_i, let us now relax the restriction $\bar{d}_{r_i}^{0i} = 0$ by setting $\bar{d}^{0i} = \gamma$ (γ some scalar) and determine the dependence of the

[1]See Section 4.1.

dual variables and reduced costs on γ. Denote these by $\lambda^*(t|\gamma)$, $\{\nu^{jr}(\gamma)\}$, $\bar{c}(t|\gamma)$ and $\{\bar{d}^{jr}(\gamma)\}$, where i here is fixed. In this analysis we shall maintain as before all the remaining relations in (9) defining λ^* and $\{\nu^{jr}\}$. Let p be such that $r_i = (\beta_i)_p$.

(85) Lemma. Let $\pi^*(\cdot)$ and $\{\mu^{jr}\}$ satisfy

$$\text{(i)}\quad \pi^*(t) = 0\,, \qquad t \in [t_i, T]$$

$$\mu^{jr} = 0\,, \qquad r > i,\ j \geq 0$$

$$\text{(ii)}\quad \pi^*(t)B._{\beta_i} + \int_t^{t_i} \pi^*(s)ds\, K._{\beta_i} + \mu^{0i}K._{\beta_i} = 0, \qquad t \in [t_{i-1}, t_i)$$

$$\mu^{0i}B._{\beta_i} + \mu^{1i}K._{\beta_i} = e_p^T$$

$$\mu^{ji}B._{\beta_i} + \mu^{j+1,i}K._{\beta_i} = 0\,, \qquad j \geq 1$$

$$\text{(iii)}\quad \pi^*(t)B._{\beta_r} + \int_t^{t_i} \pi^*(s)ds\, K._{\beta_r} + \sum_{v=r}^{i} \mu^{0v}\cdot_{\beta_r} 1_+(t_v - t) = 0,$$

$$t \in [t_{r-1}, t_r)$$

$$\mu^{jr}B._{\beta_r} + \mu^{j+1,r}K._{\beta_r} = 0\,, \qquad j \geq 0, \qquad r = 1,\ldots,i-1.$$

Then

$$\lambda^*(t|\gamma) = \lambda^*(t|0) + \gamma\pi^*(t)$$

$$\nu^{jr}(\gamma) = \nu^{jr}(0) + \gamma\mu^{jr}$$

$$\bar{c}(t|\gamma) = \bar{c}(t|0) - \gamma\{\pi^*(t)B + \int_t^T \pi^*(s)ds\, K + \sum_{v=1}^{i} \mu^{0v}K\, 1_+(t_v - t)\} \tag{86}$$

$$\bar{d}^{jr}(\gamma) = \bar{d}^{jr}(0) + \gamma(\mu^{jr}B + \mu^{j+1,r}K)\ .$$

Proof. This follows by noting that $\lambda^*(\cdot|\gamma)$ and $\{\nu^{ji}(\gamma)\}$ by definition satisfy (9) with the relation

$$\nu^{0i}B._{\beta_i} + \nu^{1i}K._{\beta_i} = 0$$

replaced by

$$\nu^{0i}(\gamma)B._{\beta_i} + \nu^{1i}(\gamma)K._{\beta_i} = \gamma e_p^T . \qquad \square$$

The relations in (86) show that the dependence of $\lambda^*(\cdot|\gamma),\ldots$ on γ takes on exceedingly simple form. Note that $\pi^*(\cdot)$ and $\{\mu^{jr}\}$ are uniquely determined and are independent of γ. Note also that by construction, $\lambda^*(\cdot|\gamma)$ and $\{\nu^{jr}(\gamma)\}$ satisfy the complementary slackness conditions for any γ. In particular, if for some γ, the triple $(\lambda^*(\cdot|\gamma), \{\nu^{jr}(\gamma)\}, \{t_r\})$ is dual feasible, then our current basic feasible solution is optimal.

Using this device of setting $\bar{d}^{0i}_{r_i} = \gamma$ and parametrizing the dual variables in terms of γ is an essential ingredient of any continuous time simplex method. This is amply demonstrated by Example E(2), where the constraints are such that there is only one feasible solution--hence trivially optimal--and yet the dual variables defined by (9) in the usual way are dual infeasible. However, parametrizing in terms of γ does yield $(\lambda^*(\cdot|\gamma), \{\nu^{jr}(\gamma)\}, \{t_r\})$ dual feasible for $\gamma \leq -1$.

Both Lehman [18] and Drews et al. [12] recognized the need for this, and suggested extending the concept of a basis. We now interpret their approach within our framework.

Suppose that we are attempting to find an improved solution using the techniques of Sections 4.2, 4.3 and 4.4, but that we get "stuck" in one of the situations mentioned at the beginning of this section. For example, suppose that at some t^*, $\bar{c}_\ell(t^*) < 0$, and that in trying to increase x_ℓ on $[t^*, t^* + \epsilon)$, we find that for all $\epsilon > 0$ we are forced to move a breakpoint t_i that fails to satisfy $p_i = p_i'$. We then parametrize the dual variables in terms of γ by relaxing the restriction $\bar{d}_{r_i}^{0i} = 0$, and thereafter fix the value of γ by imposing the restriction

$$(87) \qquad \bar{c}_\ell(t^*|\gamma) = 0 \ .$$

This corresponds in discrete time to making x_ℓ basic at the <u>point</u> t^*, and dropping x_{r_i} from the basis at the <u>point</u> t_i, with no change in the objective value. In the simplex method it is therefore a degenerate pivot whose only effect is to determine an alternative complementary set of dual variables. The same can be done in continuous time by redefining which variables are basic at certain isolated points. In [12] this is termed as a "point pair."

Since our boundedness assumption does not allow any δ's in the primal problem, assigning basic and non-basic labels to variables at isolated points can have no effect on the primal solution. However this does have a marked and important effect on the dual variables. Observe that making x_ℓ basic at t^* and dropping x_{r_i} at the

point t_i gives a count of m+1 variables basic at $t = t^*$ and m-1 basic at $t = t_i$.[1]

The next step is now to move t^* either left or right, depending on whether or not $\bar{c}_\ell^{(1)}(t^*)$ is > 0 or < 0, while simultaneously adjusting γ so as to maintain (87). The aim of doing so is to try to obtain either dual feasibility, and hence a proof of optimality, or a new interval over which we can attempt another basis change.

Whether or not a procedure of this nature will always yield either a proof of optimality, or an eventual strict decrease in the objective value is not known. It is also still an open question whether or not use of this "point pair" technique is needed <u>only</u> to obtain a proof of optimality. Posed another way, can we have a solution that is not optimal, but has the property that at no point where the dual variables, defined in the usual way by (9), are infeasible can we make a change of basis using the techniques of the previous sections and obtain a decrease in the objective value.

[1] By Theorem 2(15) we know that every right analytic extreme point has at most m variables positive over any interval. Hence having a pair of <u>intervals</u> with this surplus-shortage property is not possible.

CHAPTER 5

CONCLUSION

This dissertation has been aimed at establishing some of the fundamental concepts and results required to develop a continuous time simplex method (for the case of constant coefficients in the constraints), To this end, we have accomplished the following:

(i) A characterization of all so-called right analytic extreme points that allows us to work with them as we would with the basic feasible solutions in linear programming.

(ii) A distribution free statement of the optimality conditions.

(iii) A means of moving from one extreme point to another, with an improvement in the objective value.

Much remains unanswered from both the algorithmic and purely mathematical points of view.

On the algorithmic side there is the need to combine the constructions of Chapter 4 with an effective heuristic that chooses the intervals over which to attempt a basis change, and also decides on how large to make ϵ before recomputing the prices. Developing such a heuristic should go hand-in-hand with the design of a computer implementation. Based on small examples solved by hand, it would seem that an initial implementation should be highly interactive in nature. Of great importance in any implementation will be the question of how to solve, numerically, equations of the form

$$Dy(t) + L\int_0^t y(s)ds = d(t)$$

where the only condition on D and L is that $(\mu D + L)^{-1}$ exists for some scalar μ.

More fundamental is the question of convergence, about which nothing has been said. This is related to the (unanswered) question of when is the optimum attained at a piecewise analytic solution having only finitely many breakpoints. It is also intimately connected with the question of whether the need to take advantage of ambiguities in the dual prices, alluded to in Section 4.5, occurs only (if at all) to obtain a proof of optimality, or whether it can occur before the optimal value is reached.

The last major open question is whether or not the sufficient conditions for optimality, presented in Section 4.1, are also necessary conditions when the optimum is attained at a piecewise analytic solution having finitely many breakpoints.

APPENDIX A

RIGHT ANALYTIC FUNCTIONS

This appendix contains the propositions about right analytic functions that are required in Section 2. The main result is the proof of Lemma 2.2(7) which we restate below as Lemma (3). For completeness we review the definitions of analytic and right analytic functions.

(1) <u>Definition</u>: [22]. Let $\Omega \subset R$ be open and $g: \Omega \to R$. Then g is said to be analytic on Ω if to every open interval $I \subset \Omega$ with center a, there corresponds a series $\sum_{i=0}^{\infty} c_i(t-a)^i$ which converges to $g(t)$ for all $t \in I$.

(2) <u>Definition</u>: A function $g:[0,T] \to R$ will be called right analytic if for each $t \in [0,T)$ there is an $\epsilon > 0$ and an analytic function $h:(t-\epsilon, t+\epsilon) \to R$ such that $g(s) = h(s)$ for all $s \in [t, t+\epsilon)$.

(3) <u>Lemma</u>: Let $g:[0,T] \to R^n$ have right analytic components $g_i(\cdot)$, $i = 1,\ldots,n$. Then there exists a (possibly infinite) disjoint family of open intervals, $\{I_j\}$ such that $[0,T) = \cup_j \tilde{I}_j$,[1] and such that for each interval I_j, each g_i satisfies

(i) g_i is analytic on I_j, and

(ii) either $|g_i| > 0$ on I_j or $g_i = 0$ on I_j.

---

[1]If $I = (t',t'')$ is an open interval then $\tilde{I}$ denotes the interval $[t',t'')$.

The proof of this result will require the following lemmas.

(4) <u>Lemma</u>. Let $I \subset R$ be an open interval and $\{J_i\}$ a family of open intervals whose union is I. Let $g: I \to R$ be given. If g is analytic on each J_i then g is analytic on I.

(5) <u>Lemma</u>. Let $K \subset R$ be a compact interval, and let h be an analytic function defined on a neighborhood of K. Then either $h = 0$ on K or h has finitely many zeros in K.

(6) <u>Corollary</u>. Let $I \subset R$ be an open interval, and let $h: I \to R$ be analytic. Let $Z(h) = \{t \in I : h(t) = 0\}$. Then either $Z(h) = I$, or $Z(h)$ has no limit point in I. In the latter case $Z(h)$ is at most countable.

The proofs of both these lemmas may be found in [22].

(7) <u>Lemma</u>. Let $\{J_\alpha\}$ be any family of open intervals, and $\{I_j\}$ be a disjoint family of open intervals such that $\cup_j I_j = \cup_\alpha J_\alpha$. Then $\cup_\alpha \tilde{J}_\alpha \subseteq \cup_j \tilde{I}_j$.

<u>Proof</u>. Since the collection $\{I_j\}$ is disjoint, the connectedness of intervals implies that each J_α is contained in a unique I_j. Thus $\tilde{J}_\alpha \subseteq \tilde{I}_j$ for some j and we are done. □

Proof of Proposition (3). We shall prove the proposition for the case $n = 1$, since from this, the general case follows immediately.

By definition of g being right analytic, for each $t \in [0,T)$ there exists an $\epsilon_t > 0$ such that g is analytic on $K_t = (t, t+\epsilon_t) \subset [0,T)$. Let $V = \cup_t K_t$ and $W = \cup_t \tilde{K}_t$. Since V is open, there exists a disjoint family of open intervals $\{J_i\}_{i=1}^{\infty}$ whose union is V.

On each J_i we now obtain the following: By Lemma (4) g is analytic on J_i. By corollary (6) either $g = 0$ on J_i or $g = 0$ on at most a countable sequence $\{t_1, t_2, \ldots\} \subset J_i$. In the latter case write $J_i = (t', t'')$. Since the sequence has no limit point in J_i, and since by definition g agrees with an analytic function defined on a neighborhood of t', the only possible limit point of $\{t_k\}$ is t''. Hence we may assume that

$$t' < t_1 < t_2 < \cdots < t'' .$$

With this partition of J_i we can now conclude that there is a (possibly infinite) collection of disjoint open intervals $\{L_i^k\}$ such that $|g| > 0$ on L_i^k or $g = 0$ on L_i^k, and such that $\tilde{J}_i = \cup_k \tilde{L}_i^k$.

Now it is clear that $W = [0,T)$. By Lemma (7), it follows that $\cup_i \tilde{J}_i = [0,T)$. Hence $\cup_{i,k} \tilde{L}_i^k = [0,T)$. By relabeling the family $\{L_i^k\}$ as $\{I_j\}$, we obtain the desired result. $\square$

(7) Lemma. Let $g, h: [0,T] \to R$ be right analytic. Let $t_0 \in [0,T)$ be such that $g(t_0) \neq h(t_0)$. Then there exist $0 \leq r < s \leq T$ such that $g = h$ on $[0,r)$ and $g \neq h$ on (r,s).

Proof. Let $E = \{t: g(t) \neq h(t)\}$. Since $t_0 \in E$, E is nonempty. Let $r = \inf E$. Then by definition of r, $g = h$ on $[0,r)$. Since both g and h are right analytic there is an $\epsilon > 0$ such that g and h agree with analytic functions on $[r, r+\epsilon]$. Again by definition of r, there is a $t \in (r, r+\epsilon)$ such that $g(t) \neq h(t)$. By Lemma 5, $g - h$ has finitely many zeros on $[r, r+\epsilon]$. Hence there is an $s \in (r, r+\epsilon)$ such that $g \neq h$ on (r,s). □

APPENDIX B

EXAMPLES OF EXTREME POINTS

The constraints are all of the form

$$B(t)\ x(t) + \int_0^t K(t,s)\ x(s)ds = b(t)$$

$$x(t) \geq 0$$

for a.e. $t \in [0,T]$. $B(t)$ and $K(t,s) \in R^{m \times n}$.

(1): <u>An extreme point that is not locally uniquely defined, and is independent of small changes in</u> T.

$m = 1$, $n = 3$, $T = 5$, $B(t) = 0$ on $[0,5]$. Define $K(t,s) =$ $[\mathbf{k}_1(t,s)\ \ k_2(t,s)\ \ k_3(t,s)]$ on the triangle $0 \leq s \leq t \leq 5$ as follows:

s	t	$k_1(t,s)$	$k_2(t,s)$	$k_3(t,s)$
[0,1]	[s,1)	0	0	0
	[1,2)	e^{-st}	0	0
	[2,3)	0	e^{-st}	0
	[3,4)	0	0	e^{-st}
	[4,5]	0	0	0
(1,5]	[s,5]	0	0	0

Set

$$\bar{x}_1(t) = \bar{x}_2(t) = \bar{x}_3(t) = 1 \qquad \text{on } [0,1)$$

$$\bar{x}_1(t) = \bar{x}_2(t) = \bar{x}_3(t) = 0 \qquad \text{on } [1,5]$$

and then define $b(\cdot)$ by

$$b(t) = \int_0^t K(t,s)\,\bar{x}(s)ds, \qquad t \in [0,5]$$

Consider now the specific values of $K(t,s)$:

(i) For $t \in [0,1)$ the equation is trivial, i.e., $0 = 0$ so that the values of x on $[0,1)$ are not in any way determined by the coefficients on $[0,1)$.

(ii) For $t \in [1,2)$ the equation reads

$$\int_0^1 e^{-st}\,x_1(s)ds = b(t)\ .$$

The left-hand side is the Laplace transform of x_1 on the interval $[0,1)$, evaluated at t. By the uniqueness theorem for Laplace transforms (see Theorem C.12) the above equation has a unique (a.e.) bounded solution x_1 on $[0,1)$. By construction this solution is $x_1 = \bar{x}_1 = 1$.

(iii) For $t \in [2,3)$ and $t \in [3,4)$ we obtain similar equations in x_2 and x_3 respectively, and conclude that the only possible solution is $x_2 = \bar{x}_2$ and $x_3 = \bar{x}_3$ on $[0,1)$.

The above shows that x on $[0,1)$ is uniquely determined, independent of the choice of x on $[1,5]$. Thus choosing $x = 0$ on $[1,5]$ yields $x = \bar{x}$ on the whole interval, and this must be an extreme point solution.

Notice that we have one equality constraint in three nonnegative variables and that on $[0,1)$ all three are positive while on $[1,5]$ all three are zero. □

(2): <u>An extreme point that is locally uniquely defined except at the origin, and whose values throughout the interval depend explicitly on</u> T.

$m = n = 1$, $B(t) = t$, $b(t) = -t$ and $K(t,s) = -2$. Thus

$$tx(t) - 2 \int_0^t x(s)ds = -t$$

$$x(t) \geq 0, \qquad t \in [0,T].$$

One can easily show that the only solutions to this equation are of the form $x(t) = 1 + \alpha t$ for α an arbitrary scalar. α is the derivative of x at 0 which, once determined, yields a unique solution over the remainder of the interval. To obtain an extreme point solution we choose the least α such that $1 + \alpha t \geq 0$ on $[0,T]$. Thus the only extreme point is $x(t) = 1 - t/T$.

Note that if we set $\alpha = 0$ (say) and choose the partition $\{I_j\}$ with $I_j = (1/j+1,\ 1/j)$, then $x(t) = 1$, $t \in [0,T]$, is uniquely determined on each interval $[1/j+1,\ 1/j)$. However, it is not uniquely determined on $[0,T]$, and is also not an extreme point. $\square$

APPENDIX C

DISTRIBUTIONS, LAPLACE TRANSFORMS AND EQUATIONS OF THE FORM $Dx + L\int x = g$.

This appendix summarizes the techniques and results required to study the solutions to equations of the form

$$(1) \qquad Dx(t) + L\int_0^t x(s)ds = g(t)\,, \qquad t \geq 0$$

where $D, L \in R^{m \times m}$ and $g(\cdot)$ is analytic. We remark that by a suitable change of variable and time direction, all the results pertaining to (1) will be equally valid for equations of the form

$$(2) \qquad \lambda(t)D + \int_t^T \lambda(s)ds\, L = g(t)\,, \qquad t \leq T\,.$$

The material presented here is well known, and we refer the reader to [27], [11] and [5] for further details.

When D^{-1} exists, (1) becomes the much studied Volterra equation of the second kind. This can be shown to have the unique solution

$$(3) \qquad x(t) = D^{-1}g(t) - D^{-1}L\int_0^t e^{-D^{-1}L(t-s)}\, D^{-1}g(s)ds\,.^1$$

However, when D is singular (1) may not have a solution in any R^n-valued space of functions. Consider the following example.

---

[1] If $G \in R^{m \times m}$ then e^{Gt} denotes the matrix exponential $\sum_{n=0}^{\infty} \frac{G^n t^n}{n!}$

(4) Example.

$$D = \begin{pmatrix} 1 & 0 \\ 1 & 0 \end{pmatrix}, \quad L = \begin{pmatrix} 0 & -1 \\ 0 & 0 \end{pmatrix}, \quad g(t) = \begin{pmatrix} 0 \\ 1 \end{pmatrix}$$

i.e.

$$\begin{aligned} x_1(t) - \int_0^t x_2(s)ds &= 0 \\ x_1(t) &= 1 \end{aligned}$$

Clearly, if x_2 is restricted to be a real valued integrable function, then these equations have no solution. However we can meaningfully work within a larger class of 'functions,' the space of distributions, where these equations do have a unique solution.

C.1. Distributions

Let $\mathcal{D}$ denote the space of all C^∞ functions $\varphi: R \to R$ that vanish outside some bounded interval (depending on φ). Then the space of distributions, $\mathcal{D}'$, is defined to be the set of all continuous linear functionals on $\mathcal{D}$.[1] Thus, if $h \in \mathcal{D}'$, then to every $\varphi \in \mathcal{D}$, h assigns a real number, to be denoted by $\langle h, \varphi \rangle$. Locally integrable functions may be considered a subspace of $\mathcal{D}'$: if $f: R \to R$ satisfies

$$\int_{-M}^{M} |f(t)| dt < \infty$$

for all M, then the functional defined by

[1]See [27] for the notion of a sequence $\{\varphi_n\}$ converging in $\mathcal{D}$.

$$\langle f,\varphi\rangle = \int_{-\infty}^{\infty} f(t)\,\varphi(t)dt$$

is a continuous linear functional on $\mathcal{D}$, and hence is a member of $\mathcal{D}'$. Distributions generated this way are called regular distributions.

Of particular interest here will be the δ functional and its derivatives. δ is a distribution defined as follows:

$$\langle \delta,\varphi\rangle = \varphi(0)\ , \qquad\qquad \varphi \in \mathcal{D}\ .$$

Thus if we think of formally integrating δ against φ, $\int_{-\infty}^{\infty} \delta(t)\,\varphi(t)dt$, then δ may be thought of as a 'function' that is zero everywhere except at the origin, where it is so large that $\int_{-\infty}^{\infty} \delta(t)dt = 1$. Indeed, distributions were invented precisely to make formal notions of this kind rigorous.

Distributions allow us to differentiate functions that do not have derivatives in the usual sense. The motivation for the definition of distributional derivatives is the following. Let $f\colon R \to R$ be continuously differentiable. By definition

$$\langle f^{(1)},\varphi\rangle = \int_{-\infty}^{\infty} f^{(1)}(t)\,\varphi(t)dt$$

Integrating by parts and noting that φ vanishes outside a bounded interval, yields

$$\begin{aligned}\langle f^{(1)},\varphi\rangle &= -\int_{-\infty}^{\infty} f(t)\,\varphi^{(1)}(t)dt\\ &= -\langle f,\ \varphi^{(1)}\rangle\end{aligned}$$

Since $\varphi^{(1)}$ is also in $\mathcal{D}$, the right hand side makes sense when regarding f as a regular distribution. Thus for $\varphi \in \mathcal{D}$, $h \in \mathcal{D}'$, define $h^{(1)}$ by

$$\langle h^{(1)}, \varphi\rangle = -\langle h, \varphi^{(1)}\rangle ,$$

and more generally

$$\langle h^{(i)},\varphi\rangle = (-1)^i \langle h,\varphi^{(i)}\rangle .$$

In the case of δ, we get

$$\langle \delta^{(i)}, \varphi\rangle = (-1)^i \varphi^{(i)}(0) . \tag{5}$$

With the above notion of the δ, we can now solve Example (4) by inspection.

$$x_1(t) = 1_+(t)$$

$$x_2(t) = \delta(t)$$

What we mean by writing the solution formally in this way is that, for $\varphi \in \mathcal{D}$,

$$\langle x_1,\varphi\rangle = \int_0^\infty \varphi(t)dt$$

and

$$\langle x_2,\varphi\rangle = \varphi(0) .$$

In general, it will turn out that the solutions to (1) contain linear combinations of $\{\delta^{(i)}\}$.

An important concept in the development of solution techniques for (1) is that of <u>convolution</u>.

If $f, g: R \to R$ are two continuous functions that vanish on $(-\infty,0)$, then their convolution, $f*g: R \to R$ is defined by

$$(f*g)(t) = \int_0^t f(t-s)\, g(s)ds$$

Notice that $f*g = g*f$. This definition can be extended to the case of f and g being distributions provided $f, g \in \mathcal{D}'_+$, where $\mathcal{D}'_+ \subset \mathcal{D}'$ denotes the space of distributions 'vanishing' on $(-\infty,0)$. By this we mean that $f \in \mathcal{D}'_+$ satisfies $\langle f,\varphi\rangle = 0$ for all $\varphi \in \mathcal{D}$ vanishing on $[0,\infty)$. See [27, p. 122] for further details. Of importance here will be the result

$$\begin{aligned} \delta*f &= f \\ \delta^{(i)}*f &= f^{(i)} \end{aligned} \tag{6}$$

for $f \in \mathcal{D}'_+$.

We shall be applying (6) to functions of the form $1_+(t)f(t)$ where $f: R \to R$ is infinitely differentiable. Clearly $(1_+f)^{(i)}$ exists only in the distributional sense. It can be shown that

$$(1_+f)^{(i)} = 1_+f^{(i)} + \delta f^{(i-1)}(0) + \cdots + \delta^{(i-1)} f(0) . \tag{7}$$

C.2. Laplace Transforms

The Laplace Transform is an extremely useful tool for computing and manipulating solutions to equations of the form (1). The reason for this is that it allows us formal algebraic manipulations, the results of which are correct, usually requiring no further justification.

If $f: R \to R$ is such that $f(t) = 0$, $t < 0$, and for some $c \in R$, $e^{-ct} f(t)$ is absolutely integrable over R, then the Laplace transform of f is defined by

$$\hat{f}(s) = \int_0^\infty e^{-st} f(t)dt$$

where $s \in \mathbb{C}$, $\mathrm{Re}\, s > c$. The alternative notation $\mathcal{L}(f)(s)$ will also be used.

With some care, this definition can also be extended to distributions.

For $f \in \mathcal{D}'_+$, define

$$\hat{f}(s) = \langle f, e^{-st} \rangle$$

whenever the right-hand side makes sense. This will occur when, for some c, $e^{-ct} f(t)$ is in a certain subspace $\mathcal{S}' \subset \mathcal{D}'$ called distributions of slow growth. Any such f will be called <u>Laplace-transformable</u>. For our purpose it suffices to remark that the family $\{\delta^{(i)}\}$ is Laplace-transformable, and moreover that

$$\mathcal{L}(\delta^{(i)}) = s^i \,. \tag{8}$$

Other Laplace transform formulae that will be important are

(9) $$\mathcal{L}\left(\frac{1}{k!} t^k 1_+(t)\right) = \frac{1}{s^{k+1}}, \qquad \text{Re } s > 0$$

(10) $$\mathcal{L}\left(\frac{1}{k!} t^k e^{\mu t} 1_+(t)\right) = \frac{1}{(s-\mu)^{k+1}}, \qquad \text{Re } s > \text{Re } \mu$$

(11) $$\widehat{f*g}(s) = \hat{f}(s)\,\hat{g}(s)$$

The following uniqueness theorem gives a 1-1 correspondence between Laplace-transformable distributions and their Laplace-transforms.

(12) Theorem. (Uniqueness Theorem). Let f and g be Laplace-transformable distributions in $\mathcal{D}'_+$. If $\hat{f}(s) = \hat{g}(s)$ on some vertical line $s = c + i\omega$[1] in their regions of convergence, then $f = g$.

Proof. See [27, p. 225]. □

C.3. Solving the equation $Dx + L\int x = g$

With the above results we can now proceed to solve

$$Dx(t) + L\int_0^t x(s)ds = g(t), \qquad t \geq 0$$

[1]It suffices to require that $\hat{f}(s) = \hat{g}(s)$ over some segment of an arbitrarily oriented line in their common regions of convergence.

in the space of distributions. We begin by rewriting the equation as

$$(13) \qquad (D\delta + L1_+)*x = 1_+g, \qquad x_i \in \mathcal{D}'_+ .$$

Taking Laplace transforms on both sides (assuming x is Laplace-transformable) yields

$$(D + \frac{1}{s} L)\, \hat{x}(s) = \mathcal{L}(1_+g)(s)$$

by using (8), (9) and (11).

Assume now that $(D + \frac{1}{s} L)^{-1}$ exists for some s (and hence for all except finitely many s). Then

$$\hat{x}(s) = (D + \frac{1}{s} L)^{-1} \mathcal{L}(1_+g)(s).$$

From (11) we see that if $H = (h_{ij})$ is an $m \times m$ matrix with entries $h_{ij} \in \mathcal{D}'_+$ and whose Laplace transform is $(D + \frac{1}{s} L)^{-1}$, then $x = H*(1_+g)$.

By Cramer's rule for computing the inverse of a matrix, it is clear that $(D + \frac{1}{s} L)^{-1} = (r_{ij}(s))$ has entries that are rational functions of s. Indeed each r_{ij} has the form

$$r_{ij}(s) = \frac{p_{ij}(s)}{q(s)}$$

where the p_{ij} and q are polynomials in s of order $\leq m$, and $q(s) = \det(sD + L)$. A partial fraction expansion of each r_{ij} yields

$$(14) \qquad (D + \frac{1}{s} L)^{-1} = \sum_{k=0}^{m} U_k s^k + \sum_{k=1}^{p} \sum_{\ell=0}^{m_k-1} V_{k\ell} \frac{1}{(s-\xi_k)^{\ell+1}}$$

where $\sum_{k=1}^{p} m_k$ = degree of q, $\{\xi_k\}$ are the roots of q with multiplicities $\{m_k\}$, and the matrices $\{U_k\} \subset R^{m \times m}$, $\{V_{k\ell}\} \subset \mathbb{C}^{m \times m}$.

From (8), (10), (14) and Theorem (12) we obtain $(D + \frac{1}{s} L)^{-1}$ as the Laplace transform of

$$(15) \qquad H(t) = \sum_{k=0}^{m} U_k \delta^{(k)}(t) + \sum_{k=1}^{p} \sum_{\ell=0}^{m_k-1} V_{k\ell} \frac{1}{\ell!} t^{\ell} e^{\xi_k t} 1_+(t) .$$

To obtain the solution x, we compute the convolution $H*(1_+g)$. Using (6) and (7) yields

$$(16) \qquad x(t) = \sum_{k=0}^{m} U_k g^{(k)}(t) + \sum_{\ell=0}^{m-1} \Big(\sum_{k=\ell+1}^{m} U_k g^{(k-\ell-1)}(0) \Big) \delta^{(\ell)}(t)$$

$$+ \int_0^t \Psi(t-s)\, g(s) ds, \qquad t \geq 0$$

where

$$(17) \qquad \Psi(t) = \sum_{k=1}^{p} \sum_{\ell=0}^{m_k-1} V_{k\ell} \frac{1}{\ell!} t^{\ell} e^{\xi_k t}, ^{1} \qquad t \geq 0$$

We have thus established the following result.

---

[1] It can be shown that $\Psi(t) \in R^{m \times m}$ since the imaginary parts cancel in complex conjugate pairs.

(18) Proposition. Let D and $L \in R^{m \times m}$, and $g: R \to R^m$ be analytic. If there is a $\mu \in R$ such that $(\mu D + L)^{-1}$ exists, then the unique solution to the equation

$$Dx(t) + L \int_0^t x(s)ds = g(t), \qquad t \geq 0$$

in the space $\mathcal{D}'_+$ is given by (16).

Proof. By construction, x given by (16) is a particular solution. To show it is the unique solution in $\mathcal{D}'_+$, suppose that y is another solution. Then both x and y satisfy (13). By construction $H(t)$ given by (15) satisfies

$$H*(D\delta + L1_+) = I\delta$$

where I is the $m \times m$ identity matrix. Convolving H with both sides of (13) then yields

$$\delta * x = H*(1_+ g)$$
$$\delta * y = H*(1_+ g) .$$

By (6), it follows that $x = y$. $\square$

We remark that alternative expressions for the coefficients $\{U_k\}$ and $\Psi(t)$ have been found using the Drazin inverse of a matrix. See [5]. In particular it is shown that there exist matrices $F, G \in R^{m \times m}$ such that

$$\Psi(t) = Fe^{Gt}.$$

Since these expressions yield little additional insight into the nature of the solutions, we shall omit them here.

Finally, note that since g is assumed analytic, x is the sum of an analytic function and a finite linear combination of $\{\delta^{(i)}\}$. In the event that D^{-1} exists, it can be easily shown that $U_k = 0$, $k = 1,\ldots,m$, since q then has degree m. In this case, all the terms in $\delta^{(i)}$ in (16) vanish, and we obtain the same expression as in (3).

C.4. An initial value result

(19) Lemma. If $g: R \to R$ is analytic and $1_+ g$ is Laplace transformable then

$$\mathcal{L}\left(\sum_{k=0}^{\ell} a_k \delta^{(k)} + 1_+ g\right)(s) = \sum_{k=0}^{\ell} a_k s^k + \sum_{j=0}^{\infty} g^{(j)}(0) s^{-j}$$

Proof. Use (8), (9) and Theorem 8.3.3 in [27, p. 231]. □

APPENDIX D

AN EXISTENCE AND UNIQUENESS THEOREM RELATING TO THE $\{\tau_i\}$

In this Appendix we study the equations

$$
\begin{aligned}
& z_1 - z_2 + \cdots + (-1)^{q+1} z_q = \rho \\
& z_1^2 - z_2^2 + \cdots + (-1)^{q+1} z_q^2 = \rho \\
& \qquad \vdots \\
& z_1^q - z_2^q + \cdots + (-1)^{q+1} z_q^q = \rho
\end{aligned}
\tag{1}
$$

on the simplex

$$S_q = \{z \in R^q : 1 \geq z_1 \geq \cdots \geq z_q \geq 0\}$$

for given $0 < \rho < 1$.

These equations appear in Section 4.2.1.

We remark that they are closely related to the study of the Vandermonde matrix (see e.g. [17, p. 36])

$$V(z) = \begin{bmatrix} z_1 & z_2 & \cdots & z_q \\ z_1^2 & z_2^2 & \cdots & z_q^2 \\ \vdots & \vdots & & \\ z_1^q & z_2^q & \cdots & z_q^q \end{bmatrix}$$

al application is to solve a linear [illegible]stem of equations coefficient matrix is $V(z)$, for some [illegible]ven fixed z.

iven a solution

$$y = \begin{bmatrix} 1 \\ -1 \\ \vdots \\ (-1)^{q+1} \end{bmatrix}$$

$$V(z)y = \rho e$$

e equation

d wish to determine z.

The main result of this Appendix is the following:

(2) Theorem. The system of equations (1) h[illegible]que solution [illegible] the interior of the simplex S_q.

We shall require severa[illegible]sults for the proof of this the[illegible].

(3) Lemma. For an[illegible] and any $z \in R^q$ satisfying (1),

(4) $[illegible] - z_2^{q+1} + \dots + (-1)^{q+1} z_q^{q+1} = \rho(1 - \prod_{i=1}^{q} (1-z_i))$.

[illegible]oof. Define the polynomial

$$f(\omega) = \omega \prod_{i=1}^{q} (\omega - z_i)$$

and let it have the expansion

However, the usual application is to solve a linear system of equations whose detached coefficient matrix is $V(z)$, for some given fixed z. Here we are given a solution

$$y = \begin{bmatrix} 1 \\ -1 \\ \vdots \\ (-1)^{q+1} \end{bmatrix}$$

to the equation

$$V(z)y = \rho e$$

and wish to determine z.

The main result of this Appendix is the following:

(2) <u>Theorem</u>. The system of equations (1) has a unique solution in the interior of the simplex S_q.

We shall require several results for the proof of this theorem.

(3) <u>Lemma</u>. For any ρ and any $z \in R^q$ satisfying (1),

(4) $$z_1^{q+1} - z_2^{q+1} + \cdots + (-1)^{q+1} z_q^{q+1} = \rho\{1 - \prod_{i=1}^{q} (1-z_i)\}.$$

<u>Proof</u>. Define the polynomial

$$f(\omega) = \omega \prod_{i=1}^{q} (\omega - z_i)$$

and let it have the expansion

$$f(\omega) = \omega^{q+1} + a_q\omega^q + \cdots + a_1\omega .$$

Multiply the ith equation in (1) by a_i and sum over $i = 1,\ldots,q$. It is easily seen that the resulting equation is

$$\{f(z_1) - z_1^{q+1}\} + \cdots + (-1)^{q+1}\{f(z_q) - z_q^{q+1}\} = \rho\{f(1) - 1\}.$$

Since $f(z_i) = 0$ and $f(1) = \prod_{i=1}^{q} (1 - z_i)$, the result follows. □

(6) Corollary. For $0 < \rho < 1$, the $(q+1)$ equations

$$\sum_{k=1}^{q} (-1)^{k+1} z_k^j = \rho , \qquad j = 1,2,\ldots,q+1$$

have no solution in the interior of S_q.

Proof. For z in the interior of S_q, the right-hand side of (4) is $< \rho$, and the result follows immediately. □

(7) Corollary. For $0 < \rho < 1$, equations (1) have no solution on the boundary of S_q.

Proof. Let z be the boundary of S_q and satisfy (1). By examining the form of such z and the form of (1), it is clear that there are $1 \leq \ell_1 < \cdots < \ell_r \leq q$ such that $(z_{\ell_1},\ldots,z_{\ell_r})$ is in the interior of S_r and satisfies

$$\sum_{i=1}^{r} (-1)^{i+1} z_{\ell_i}^{j} = \rho' , \qquad j = 1,2,\ldots,r+1$$

where either $\rho' = \rho$ or $\rho' = 1-\rho$. By Corollary (6) this is a contradiction. □

(8) Lemma. Det $V(z) = \prod_{1 \le k \le q} z_k \prod_{1 \le i < j \le q} (z_j - z_i)$.

Proof. See [17, p. 36]. □

(9) Corollary. Define $f: R^q \to R^q$ by

$$f_j(z) = \sum_{k=1}^{q} (-1)^{k+1} z_k^j , \qquad j = 1,\ldots,q$$

Then

$$\det Df(z) = q! \, (-1)^{q+1} \prod_{1 \le i < j \le q} (z_j - z_i)$$

where $Df(z)$ is the Jacobian matrix of $f(z)$.

Proof. This follows directly from the Lemma and elementary properties of determinants. □

(10) <u>Lemma</u>. Let $\gamma_1,..,\gamma_q \in R$ be given. If the equation

$$\begin{aligned} y_1 + y_2 + \cdots + y_q &= \gamma_1 \\ y_1^2 + y_2^2 + \cdots + y_q^2 &= \gamma_2 \\ &\vdots \\ y_1^q + y_2^q + \cdots + y_q^q &= \gamma_q \end{aligned}$$

have a solution, then this solution is unique up to a permutation of the variables.

<u>Proof</u>. This follows from [4, pp. 242-245]. □

<u>Proof of Theorem (2)</u>. <u>Existence</u>: The proof will be by induction on q.

For the case $q = 1$ the result holds trivally.

Suppose that the result is true for q, and that $\bar{z}$, in the interior of S_q, is a solution. Define the maps $H: S_q \times [0,1] \to R^q$ and $h: S_q \times [0,1] \to R$ by

$$H_j(z,t) = \sum_{k=1}^{q} (-1)^{k+1} z_k^j + (-1)^q t^j \, , \qquad j = 1,\ldots,q$$

and

$$h(z,t) = \sum_{k=1}^{q} (-1)^{k+1} z_k^{q+1} + (-1)^q t^{q+1}$$

We need to find $1 > z_1 > \cdots > z_q > t > 0$ such that $H(z,t) = \rho e$ and $h(z,t) = \rho$.

Now $H(\bar{z},0) = \rho e$, and by Lemma (3), $h(\bar{z},0) < \rho$ since $\bar{z}$ is in the interior of S_q. Also by Corollary (9) the Jacobian, $D_z H(z,t)$ is nonsingular for all z in the interior of S_q and all t. By the Implicit Function Theorem [20, p. 128] there is a $\delta > 0$ such that for all $t \in [0,\delta)$, the equation $H(z,t) = \rho e$ has a solution $z(t)$ in the interior of S_q. Moreover $z(t)$ varies continuously with t.

We now show that we can increase t from 0 to some t^* while maintaining $H(z(t),t) = \rho e$, $z(t)$ in the interior of S_q, $0 < t < z_q(t)$, and such that $h(z(t^*),t^*) = \rho$. If this is possible then $(z(t^*),t^*)$ will be in the interior of S_{q+1} and will satisfy (1) in the case of q replaced by $q+1$.

Since $D_z H(z,t)$ is nonsingular for all z in the interior of S_q, it is clear that the only time we can no longer increase t is when either $t = z_q(t)$ or $z(t)$ hits the boundary of S_q. In the event that $t = z_q(t)$ or any two components of $z(t)$ are equal, we can use the technique of the proof of Corollary (7) to obtain a reduced overdetermined system having an interior point solution. By Corollary (6) this leads to a contradiction. The case remaining is when for some t', $z_1(t') = 1$ and $1 > z_2(t') > \cdots > z_q(t') > t' > 0$. In this case $(z_2(t'), \ldots, z_q(t'),t')$ satisfies (1) with the right-hand side replaced by $1 - \rho$. By Lemma (3) we get

$$\sum_{k=2}^{q} (-1)^k z_k^{q+1} + (-1)^{q+1} t'^{q+1}$$
$$= (1-\rho) \{1 - (1-t') \prod_{k=2}^{q} (1 - z_k)\} .$$

Hence

$$h(z(t'),t') = 1 - (1-\rho) \{1 - (1-t') \prod_{k=2}^{q} (1-z_k)\} .$$

Since the term { } lies in (0,1) we have

$$h(z(t'),t') > \rho .$$

However, $h(z(0),0) < \rho$, and we would have reached an earlier $t^* < t'$ at which

$$h(z(t^*),t^*) = \rho.$$

This completes the proof that the result holds for q+1. By induction the existence is established.

<u>Uniqueness</u>: (This proof is due to T. J. Rivlin.[1]) Suppose there are two solutions z and $\bar z$. Upon rearrangement, it is clear that z and $\bar z$ satisfy

$$\begin{aligned}
z_1 + \bar z_2 + z_3 + \cdots &= \bar z_1 + z_2 + \bar z_3 + \cdots \\
z_1^2 + \bar z_2^2 + z_3^2 + \cdots &= \bar z_1^2 + z_2^2 + \bar z_2^2 + \cdots \\
&\vdots \\
z_1^q + \bar z_2^q + z_3^q + \cdots &= \bar z_1^q + z_2^q + \bar z_3^q + \cdots .
\end{aligned}$$

By Lemma (10) the vector $(z_1,\bar z_2,z_3,\ldots)$ is a permutation of $(\bar z_1,z_2,\bar z_3,\ldots)$. Since both z and $\bar z$ lie in the interior of S_q, it follows that $z_i = \bar z_i$, $i = 1,\ldots,q$.

This completes the proof. □

---

[1]Private communication.

APPENDIX E

EXAMPLES OF SOME STEPS IN THE CONTINUOUS TIME SIMPLEX ALGORITHM

(1). Making a change of basis over $[0,\epsilon)$ with $q = 2$.

$$x_1(t) + \int_0^t x_2(s)ds = 2 + 2t$$

$$x_2(t) + \int_0^t x_3(s)ds = 2 + t$$

$$x_3 + x_5 = 2$$

$$x_1 + x_4 = 2$$

$$x_i(t) \geq 0, \quad i = 1,\ldots,5, \ t \in [0,1] .$$

Since we are using this example only to illustrate the procedure for making a change of basis over $[0,\epsilon)$, we have omitted the objective function.

The detached coefficients are:

$$B = \begin{bmatrix} 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 1 & 0 & 0 & 1 & 0 \end{bmatrix}, \qquad K = \begin{bmatrix} 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}$$

$$b(t) = \begin{bmatrix} 2 + 2t \\ 2 + t \\ 2 \\ 2 \end{bmatrix} .$$

We begin with the initial basic feasible solution $\beta_1 = (1,2,3,5)$ over $[0,1]$. This yields

$$x_1(t) = x_2(t) = 2, \quad x_3(t) = x_5(t) = 1, \quad x_4(t) = 0 .$$

Next we introduce $x_4 = \theta$ into the basis over $[0,\epsilon)$. Following Section 4.2.1 the first step is to find the partial fraction expansion of

$$\hat{h}(s) = (B_{\cdot\beta_1} + \frac{1}{s} K_{\cdot\beta_1})^{-1} (b_4 + \frac{1}{s} k_4) .$$

Substituting yields

$$\hat{h}(s) = \begin{bmatrix} 1 & 1/s & 0 & 0 \\ 0 & 1 & 1/s & 0 \\ 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 \end{bmatrix}^{-1} \begin{bmatrix} 0 \\ 0 \\ 0 \\ 1 \end{bmatrix} = \begin{bmatrix} 1 \\ -s \\ s^2 \\ -s^2 \end{bmatrix}$$

whose expansion is

$$\hat{h}(s) = \begin{bmatrix} 0 \\ 0 \\ 1 \\ -1 \end{bmatrix} s^2 + \begin{bmatrix} 0 \\ -1 \\ 0 \\ 0 \end{bmatrix} s + \begin{bmatrix} 1 \\ 0 \\ 0 \\ 0 \end{bmatrix} .$$

Thus $q = 2$.

Following (21), (29) and (33) of Section 4.2.1 we find r_1, ρ_1, r_2, ρ_2 by taking minimum and maximum ratios on the vectors

$$x(0) = (2, 2, 1, 0, 1)^T$$
$$y = (0, 0, 1, 0, -1)^T$$

We obtain

$$r_1 = 3, \quad \rho_1 = 1, \quad r_2 = 5, \quad \rho_2 = 1.$$

Since $q = 2$, we partition the interval $[0,\epsilon)$ into three subintervals $[0,\tau_1\epsilon)$, $[\tau_1\epsilon, \tau_2\epsilon)$, and $[\tau_2\epsilon,\epsilon)$ on which we shall have the basic sets $\sigma_1, \sigma_2, \sigma_1$ respectively, where

$$\sigma_1 = (1,2,4,5)$$
$$\sigma_2 = (1,2,3,4) .$$

To obtain τ_1 and τ_2 we begin by solving the equations 4.2.1 (36) for z_1 and z_2, and then set $\tau_1^0 = 1 - z_1$, $\tau_2^0 = 1 - z_2$ as our initial guess. Should we obtain

$$\theta(\epsilon | \tau_1^0, \tau_2^0) = 0$$
$$\theta^{(1)}(\epsilon | \tau_1^0, \tau_2^0) = 0$$

then these will be the desired values of τ_1 and τ_2. Now z_1 and z_2 satisfy

$$z_1 - z_2 = 1/2$$
$$z_1^2 - z_2^2 = 1/2$$
$$1 > z_1 > z_2 > 0 .$$

Solving yields

$$z_1 = 3/4, \quad z_2 = 1/4 ,$$

and therefore

$$\tau_1^0 = 1/4, \quad \tau_2^0 = 3/4 .$$

Solving the equations with these values for τ_1 and τ_2 shows that θ does satisfy the above conditions at $t = \epsilon^-$. The solution is given as follows:

	$[0, \frac{1}{4}\epsilon)$	$[\frac{1}{4}\epsilon, \frac{3}{4}\epsilon)$	$[\frac{3}{4}\epsilon, \epsilon)$	$[\epsilon, 1)$
x_1	$2 - \frac{1}{2}t^2$	$2 - \frac{1}{2}t^2 + (t - \frac{1}{4}\epsilon)^2$	$2 - \frac{1}{2}t^2 + (t - \frac{1}{4}\epsilon)^2 - (t - \frac{3}{4}\epsilon)^2$	2
x_2	$2 + t$	$2 + \frac{1}{2}\epsilon - t$	$2 - \epsilon + t$	2
x_3	0	2	0	1
$x_4 = \theta$	$\frac{1}{2}t^2$	$\frac{1}{2}t^2 - (t - \frac{1}{4}\epsilon)^2$	$\frac{1}{2}t^2 - (t - \frac{1}{4}\epsilon)^2 + (t - \frac{3}{4}\epsilon)^2$	0
x_5	0	0	2	1

Below we sketch the solution.

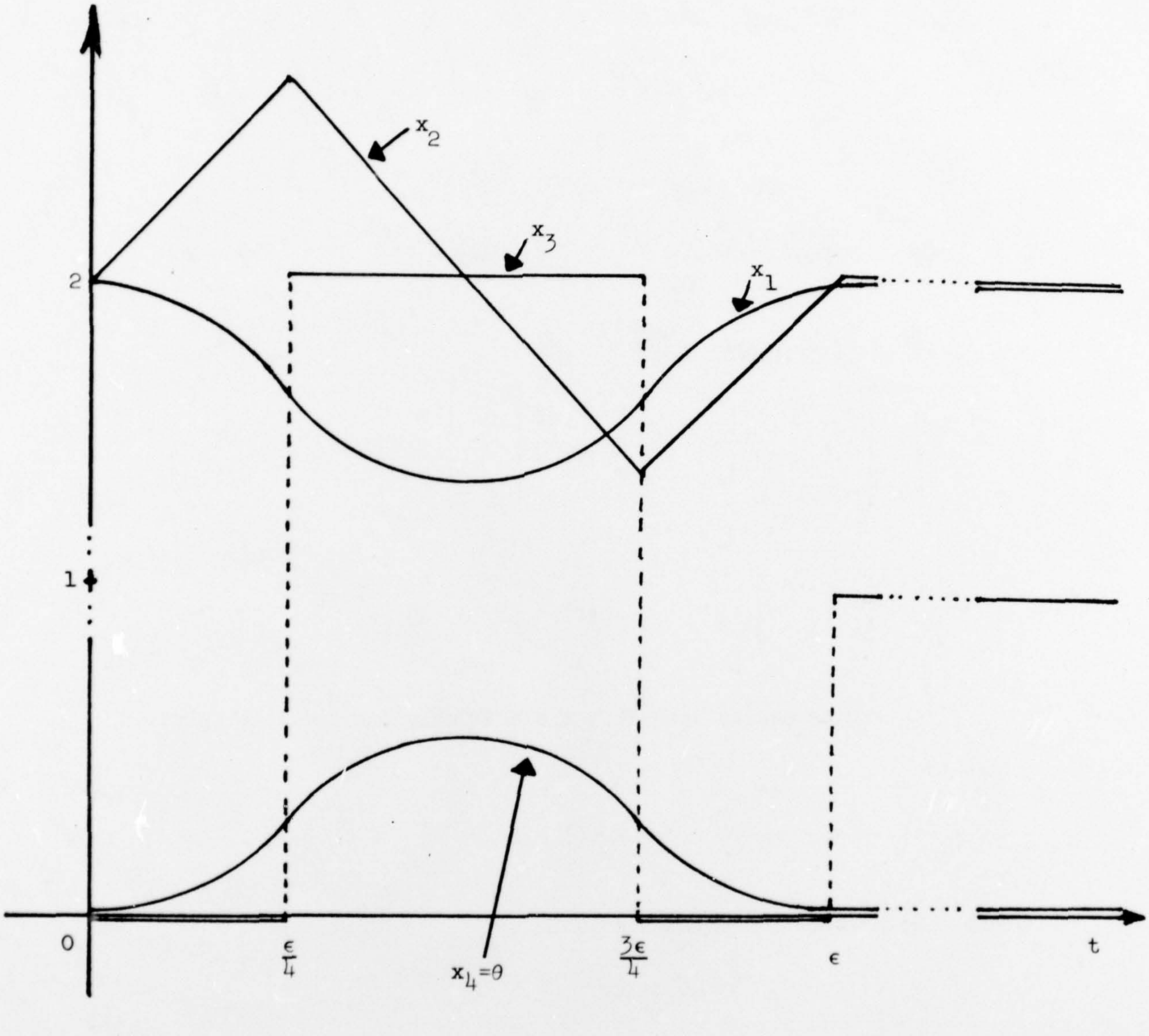


FIGURE 10

(2). Using ambiguity in the dual variables to obtain a proof of optimality

$$\text{minimize} \quad \int_0^1 x_1(t)dt$$

$$\text{subject to} \quad x_1(t) + \int_0^t x_2(s)ds = 1 - t$$

$$x_2(t) + x_3(t) = 1$$

$$x_i(t) \geq 0 , \qquad t \in [0,1]$$

The detached coefficients are

$$B = \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 1 \end{bmatrix} , \qquad K = \begin{bmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \end{bmatrix}$$

$$b(t) = \begin{bmatrix} 1 - t \\ 1 \end{bmatrix} \qquad c(t) = [1 \;\; 0 \;\; 0] .$$

It can be easily seen that there is only one feasible solution, given by

$$x_1(t) = 1\text{-}t, \;\; x_2(t) = 0, \;\; x_3(t) = 1 , \qquad t \in [0,1] .$$

Hence this must be the optimal solution.

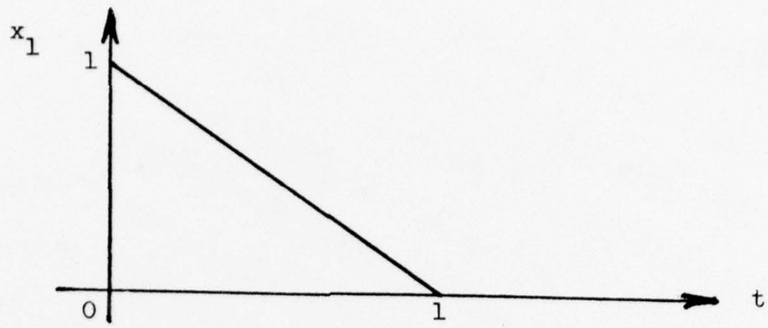


FIGURE 11

Determining the dual variables by solving 4.1(9) in the usual way yields

$$\lambda_1^*(t) = 1, \qquad \lambda_2^*(t) = 0$$

$$\nu_r^j = 0 \,, \quad r = 1,2, \quad j = 0,1 \,.$$

The reduced costs of x_2 are

$$\bar{c}_2(t) = t - 1$$

$$\bar{d}_2^j = 0 \,, \qquad j = 0,1$$

and we observe that $\bar{c}_2(t) < 0$ for $0 \leq t < 1$.

Noting that $x_1(1^-) = 0$, we can relax the restriction $\bar{d}_1^0 = 0$, and parametrize the dual variables in terms of $\gamma = \bar{d}_1^0$. Solving the equations given by Lemma 4.5(85):

$$(\pi_1^*, \pi_2^*) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + \int_t^1 (\pi_1^*, \pi_2^*) ds \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} + (\mu_1^0, \mu_2^0) \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = (0,0)$$

$$(\mu_1^0, \mu_2^0) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} + (\mu_1^1, \mu_2^1) \begin{pmatrix} 0 & 0 \\ 0 & 0 \end{pmatrix} = (1,0)$$

$$(\mu_1^1, \mu_2^1) \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix} = (0,0),$$

yields

$$\pi_1^*(t) = \pi_2^*(t) = 0$$

$$\mu_1^0 = 1, \text{ all other } \mu_r^j = 0 .$$

Thus

$$\lambda_1^*(t|\gamma) = 1, \qquad \lambda_2^*(t|\gamma) = 0$$

$$\nu_1^0(\gamma) = \gamma, \quad \text{all other} \quad \nu_r^j(\gamma) = 0.$$

The reduced costs of x_2 now become

$$\bar{c}_2(t|\gamma) = t - 1 - \gamma$$

$$\bar{d}_r^j = 0, \quad \text{all } r, j \quad .$$

We see that $\bar{c}_2(t|\gamma) \geq 0$ for $\gamma \leq -1$. Since this also yields $\bar{d}_1^0 = \gamma < 0$ for such γ, our dual variables $\lambda^*(\cdot|\gamma)$ and $\{\nu^j(\gamma)\}$ are dual feasible in this range, as required.

BIBLIOGRAPHY

[1] Beale, E.M.L., "Sparseness in Linear Programming," in Large Sparse Sets of Linear Equations, J. K. Reid (ed.) (1971), Academic Press, London, pp. 1-15.

[2] Bellman, R.E., "Bottleneck Problems and Dynamic Programming," Proc. Nat. Acad. Sci. Vol 39 (1953).

[3] Bellman, R.E., Dynamic Programming, Princeton (1972).

[4] Bôcher, M., Introduction to Higher Algebra, MacMillan, New York (1907).

[5] Campbell, S.L., Meyer, C.D. Jr., and Rose, N.J., "Applications of the Drazin Inverse to Linear Systems of Differential Equations with Singular Constant Coefficients," SIAM J. Applied Math., Vol. 31, No.3 November 1976.

[6] Dantzig, G.B., Linear Programming and Extensions, Princeton University Press, Princeton, New Jersey (1963).

[7] Dantzig, G.B., "Large Scale Systems Optimization with Applications to Energy," Technical Report SOL 77-3, April 1977, Department of Operations Research, Stanford University, Stanford, California.

[8] Dantzig, G.B., "Upper Bounds, Secondary Constraints, and Block Triangularity in Linear Programming," Econometrica Vol. 23, April 1955, pp. 174-183.

[9] Dantzig, G.B., "Linear Control Processes and Mathematical Programming," SIAM Journal of Control, Series A, 4 (1), 1966, pp. 56-60.

[10] Dantzig, G.B., unpublished notes.

[11] Doležal, V. Dynamics of Linear Systems, Academia, Prague (1967).

[12] Drews, W.P., Hartberger, R.J. and Segers, R.B., "On Continuous Mathematical Programming," in Optimization Methods in Resource Allocation, edited by R.W. Cottle and J. Kraup, Crane,,Russak and Co. Inc., N.Y., 1974.

[13] Dubovitskii, A.Ya. and Milyutin, A.A., "Necessary Conditions for a Weak Extremum in Optimal Control Problems with Mixed Constraints of the Inequality Type," USSR Computational Mathematics and Mathematical Physics 8, No. 4, pp. 725-779 (1968).

[14] Grinold, R.C., "Continuous Programming, Part One: Linear Objectives," J. Math. Analysis and Applications, Vol. 28, No. 1, October 1969, pp. 32-51.

[15] Hager, W.W. and Mitter, S.K., "Lagrange Duality Theory for Convex Control Problems," J. Control and Optimization, Vol. 14 No. 5, August 1976, pp. 843-856.

[16] Holmes, R.B., Geometric Functional Analysis and its Applications, Springer-Verlag, 1975.

[17] Knuth, D.E., The Art of Computer Programming, Volume 1: Fundamental Algorithms, Addison-Wesley, 1968.

[18] Lehman, R.S., "On the Continuous Simplex Method," RAND Research Memorandum RM-1386, Santa Monica, California (1954).

[19] Levine, P. and Pomerol, J.C., "C-closed Mappings and Kuhn-Tucker Vectors in Convex Programming," Discussion paper 7620, Center for Operations Research and Economics, Universite Catholique de Louvain, Heverlee, Belgium (1976).

[20] Ortega, J.M. and Rheinboldt, W.C., Iterative Solution of Nonlinear Equations in Several Variables, Academic Press, 1970.

[21] Perold, A.F. and Dantzig, G.B., "A Basis Factorization Method for Block Triangular Linear Programs," Technical Report SOL 78-7, April 1978, Department of Operations Research, Stanford University, Stanford, California.

[22] Rudin, W., Real and Complex Analysis, McGraw-Hill, 1974.

[23] Silverman, L.M., "Inversion of Multivariable Linear Systems," IEEE Transactions on Automatic Control, Vol. AC-14, No. 3, June 1969.

[24] Teren, F., "Minimum Time Acceleration of Aircraft Turbofan Engines by Using an Algorithm Based on Nonlinear Programming," NASA Technical Memorandum TM-73741 Lewis Research Center, Cleveland, Ohio, September 1977.

[25] Tricomi, F.G., Integral Equations, Interscience, 1957.

[26] Tyndall, W.F., "A Duality Theorem for a Class of Continuous Linear Programming Problems," SIAM J. Applied Mathematics (1965 , pp. 644-666.

[27] Zemanian, A.H., Distribution Theory and Transform Analysis, McGraw-Hill, 1965.

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER SOL 78-26	2. GOVT ACCESSION NO.	3. RECIPIENT'S CATALOG NUMBER
4. TITLE (and Subtitle) FUNDAMENTALS OF A CONTINUOUS TIME SIMPLEX METHOD		5. TYPE OF REPORT & PERIOD COVERED Technical Report
		6. PERFORMING ORG. REPORT NUMBER SOL 78-26
7. AUTHOR(s) André F. Perold		8. CONTRACT OR GRANT NUMBER(s) N00014-75-C-0267
9. PERFORMING ORGANIZATION NAME AND ADDRESS Department of Operations Research -- SOL Stanford University Stanford, CA 94305		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS NR-047-143
11. CONTROLLING OFFICE NAME AND ADDRESS Operations Research Program -- ONR Department of the Navy 800 N. Quincy Street, Arlington, VA 22217		12. REPORT DATE December 1978
		13. NUMBER OF PAGES 147
14. MONITORING AGENCY NAME & ADDRESS(if different from Controlling Office)		15. SECURITY CLASS. (of this report) UNCLASSIFIED
		15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

This document has been approved for public release and sale; its distribution is unlimited.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Continuous linear programming
Optimal control
Mixed constraints
Extreme Points
Infinite dimensional optimization
Basic feasible solutions
Continuous time simplex method
Duality
Optimal conditions

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

Continuous linear programs have wide applicability as models of many real world situations that are intertemporal in nature. An important special case is the all linear optimal control problem with mixed state and control variable constraints. This dissertation considers generalizations of the techniques of linear programming to solve continuous linear programs with constant coefficients. A characterization of extreme points is obtained in terms of certain full rank conditions. Such solutions are called "basic → next page

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE
S/N 0102-014-6601

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

delta

20. Abstract continued

feasible solutions. Then a statement of the optimality conditions is presented in a framework that is free of the δ functional and its derivatives. This is followed by the major part of the dissertation which is a means of obtaining descent by moving along a continuous path of extreme points. This work is presented in the same spirit as that of Lehman (1954) and Drews, Hartberger and Segers (1970).